NARA Regulatory Science Course Readings

NARA Regulatory Science Course Readings

This book of readings will provide the learner in the course with all the basics to the theory, paradigm, measurement, protocols and potential outcomes related to regulatory compliance within the overall field of regulatory science.

The readings are organized in the following manner:

Measurement: Uncertainty-Certainty Matrix

Overview to the theory, paradigm, approaches and methodologies: American Scientist

TRC: Theory of Regulatory Compliance

TRC+: The paradigms and program monitoring systems

Model: Early Childhood Program Quality Improvement and Indicator Model

Conceptual Framework: Differential monitoring, risk assessment, key indicators

Risk Assessment: Stepping Stones

Risk Assessment and Key Indicators: Caring for Our Children Basics History

Key Indicators: Parents Guide to Health and Safe Child Care

Key Indicators: Program quality indicators, the Saskatchewan Project and Study

Professional Development and Coaching: CYFC article on Mentoring

Coaching and Caring for Our Children: CCHC article

Effectiveness of Online Training: Early Childhood Education article

Effectiveness of Online Training: Better Kid Care evaluation report

Other Quality Initiatives: NECPA accreditation system evaluation

Scaling and Measurement: Regulatory Compliance Scale

knowledge

MDPI

The Uncertainty-Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies

Richard Fiene

Edna Bennett Pierce Prevention Research Center,
The Pennsylvania State University

Correspondence: rfiene@rikinstitute.com

Abstract: This research article will propose the use of an uncertainty-certainty matrix for licensing decision making in the human services. It will show how the matrix can be used in rule decision making and how it clearly shows when decision making has gone awry. It is also used in making decisions in differential monitoring and in validation and reliability studies.

Keywords: Decision Making; Uncertainty; Regulatory Compliance; Licensing.

Introduction

This research abstract will take the Confusion Matrix which is a well-known metric in the decision-making research literature and refocus it on regulatory science within the context of the definition of regulatory compliance and licensing measurement. It will also deal with the policy implications of this particular metric. In this abstract, it is proposed that the Uncertainty-Certainty Matrix (UCM) is a fundamental building block to licensing decision making. The 2 x 2 matrix has been written about in several posts in this blog and is the center piece for determining key indicator rules, but it is also a core conceptual framework in licensing measurement and ultimately in program monitoring and reviews.

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

The reason for selecting this matrix is the nature of licensing data, it is binary or nominal in measurement. Either a rule/regulation is in compliance or out of compliance. Presently most jurisdictions deal with regulatory compliance measurement in this nominal level or binary level. There is to be no gray area, this is a clear distinction in making a licensing decision about regulatory compliance. The UCM also takes the concept of Inter-Rater Reliability (IRR) a step further in introducing an uncertainty dimension that is very important in licensing decision making which

is not as critical when calculating IRR. It is moving from an individual metric to a group metric (See Figures 1 & 2) involving regulatory compliance with rules.

The key pieces to the UCM are the following: the decision (D) regarding regulatory compliance and actual state (S) of regulatory compliance. Plus (+) = In-compliance or Minus (-) = Out of compliance. So, let's build the matrix:

Table 1: Uncertainty-Certainty Matrix (UCM) Logic Model

UCM Matrix Logic		**Decision (D) Regarding**	**Regulatory Compliance**
		(+) In Compliance	(-) Not In Compliance
Actual State (S) of	(+) In Compliance	Agreement	Disagreement
Compliance	(-) Not In Compliance	Disagreement	Agreement

The above UCM matrix demonstrates when agreement and disagreement occur which establishes a level of certainty (Agreement Cells) or uncertainty (Disagreement Cells). In a perfect world, there would only be agreements and no disagreements between the decisions made about regulatory compliance and the actual state of regulatory compliance. But from experience, this is not the case based upon reliability testing done in the licensing research field in which a decision is made regarding regulatory compliance with a specific rule or regulation and then that is verified by a second observer who generally is considered the measurement standard.

Disagreements raise concerns in general, but the disagreements are of two types: false positives and false negatives. A false positive is when a decision is made that a rule/regulation is out of compliance when it is in compliance. Not a good thing but its twin disagreement is worse where with false negatives it is decided that a rule/regulation is in compliance when it is out of compliance. False negatives need to be avoided because they place clients at extreme risk, more so than a false positive. False positives should also be avoided but it is more important to deal with the false negatives first before addressing the false positives.

Methods

Let's look at this from a mathematical point of view in the following matrix. In order to better understand the above relationships and determine when ameliorative action needs to occur to shore up the differences between the agreements and disagreements, it is easier to do this mathematically than trying to eyeball it.

Table 2: Uncertainty-Certainty Matrix (UCM) Math Model

UCM Matrix Math Model		**Decision (D) Regarding**	**Regulatory Compliance**	**Totals**
		(+) In Compliance	(-) Not In Compliance	
Actual State (S)	(+) In Compliance	A	B	Y
Of Compliance	(-) Not In Compliance	C	D	Z
Totals		W	X	

Formulae based upon above:

Agreements = (A)(D); Disagreements = (B)(C); Randomness = sqrt ((W)(X)(Y)(Z))

UCM Coefficient = ((A)(D)) - ((B)(C)) / sqrt ((W)(X)(Y)(Z)) in which a coefficient closer to 1 indicates agreement (certainty) and a coefficient closer to –1 indicates disagreement (uncertainty). A coefficient closer to 0 indicates randomness. Obviously, we want to see (A)(D) being predominant and very little in (B)(C) which are false positives and negatives where decisions and the actual state of regulatory compliance are not matching. If (WXYZ) is predominant then there is just randomness in the data. Also, not an intended result.

The reason for even suggesting this matrix is the high level of dissatisfaction with the levels of reliability in the results of program monitoring reviews as suggested earlier. If it were not so high, it would not be an issue; but with it being so high the field of licensing needs to take a proactive role in determining the best possible way to deal with increasing inter-rater reliability among licensing inspectors. Hopefully, this organizational schema via the UCM Matrix will help to think through this process related to licensing measurement and monitoring systems.

$$\boldsymbol{UCM} = \ll \boldsymbol{A} \times \boldsymbol{D} \gg - \ll \boldsymbol{B} \times \boldsymbol{C} \gg \ \div \sqrt{\ } \ll \boldsymbol{W} \times \boldsymbol{X} \times \boldsymbol{Y} \times \boldsymbol{Z} \gg$$

The above formula provides a means to calculate when action needs to be taken based upon the respective UCM coefficients. A UCM coefficient from +.25 to +1.00 is in the acceptable range; +.24 to -.24 is due to randomness and needs to be addressed with additional inter-rater reliability training; -.25 to –1.00 indicates a severe disagreement problem that needs to be addressed both in reliability training and a full review of the targeted rules/regulations to determine if the specific rule needs additional clarification.

Table 3: Uncertainty-Certainty Matrix (UCM) Licensing Decision Coefficient Ranges

UCM Coefficient	**Licensing Decision**
+.25 to +1.00	Acceptable, No Action Needed, In or Out of Regulatory Compliance Verified through mostly Agreements. (Generally, 90% of cases)
+.24 to -.24	Random, Agreements + Disagreements, Needs Reliability Training. (Generally, 5% of cases)
-.25 to –1.00	Unacceptable, Mostly Disagreements, Needs Training & Rule/Regulation Revision. (Generally, 5% of cases)

Figure 1: Kappa Coefficient

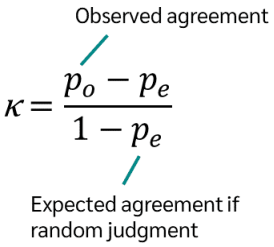


Figure 2: Uncertainty-Certainty Coefficient

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Let's provide an example of how this could work. A standard/rule/regulation that is common is the following:

Do all caregivers/teachers and children wash their hands often, especially before eating and after using the bathroom or changing diapers?

This is obviously an observation item where the licensing staff would observe in a sample of classrooms in a child care center for a set period of time. During their observations, there were several opportunities where the necessary behavior was required, and the staff complied with the rule and washed their hands. So, on the surface this specific rule was in compliance and there would appear to be full compliance with this rule based upon the observation.

A second scenario is where the observation is made, and the licensing staff observes the child care staff not washing their hands on several occasions. Then this specific rule would be out of compliance, and it would be duly noted by the licensing staff. These two scenarios establish a certain level of certainty during this observation session. However, there are other outcomes, for example, possibly one of the classrooms that was not observed had the opposite finding than what was observed in these particular classrooms. If data were being aggregated and a specific percentage was to be used the final decision about this rule could be different. Now we are getting into the uncertainty cells of the matrix where a false positive or negative could be the result. The licensing staff records the rule as being in compliance when in reality it is not = false negative or the rule is recorded as being out of compliance when in reality it is in compliance = false positive.

Another example which involves either Random Clinical Trials (RCT) or the use of abbreviated inspections (AI) and the results from these two interventions. The decision making in both RCT and AI is

basically the same. We want to make sure that the results match reality. Every time an abbreviated review is done the following four regulatory compliance results should occur based upon the UCM matrix: 1) no additional random non-compliance is found; 2) there are no false negatives (abbreviated review finds no non-compliance but in reality there is); 3) when there is non-compliance found in abbreviated inspections, other related non-compliance is found; and 4) lastly the level of false positives (abbreviated review finds non-compliance but in reality there are no other related non-compliances) is kept to a minimum. This last result based upon copious research is that it is difficult to obtain but as the regulatory science moves forward hopefully this will become more manageable.

Hopefully these above examples provided some context for how the Uncertainty-Certainty Matrix (UCM) can be used in making specific licensing decisions based upon the regulatory compliance results.

Results

Uncertainty-Certainty Matrix for Validation and Reliability Studies

The purpose of this part of this research abstract is to explore the possibility of utilizing the Uncertainty-Certainty Matrix (UCM) in validation and reliability studies in licensing decision making. The UCM has been proposed for use in licensing decision making but this would be an extension of this thinking to studies that involve validating licensing decisions such as when key indicators are used in comparison with comprehensive reviews of rules, and in reliability studies to determine individual inspector bias in regulatory compliance.

The basic premise of the UCM is that individual decision-making matches reality. When it comes to regulatory compliance decision making a 2 x 2 matrix can be drawn with the possible outcomes as is indicated in the following table (Table 4).

Table 4

UCM Matrix Logic		**Decision Regarding**	**Regulatory Compliance**
		(+) In Compliance	(-) Not In Compliance
Actual State of	(+) In Compliance	Agreement (++)	Disagreement (+-)
Compliance	(-) Not In Compliance	Disagreement (-+)	Agreement (--)

In using this table, the hope is that the decision regarding regulatory compliance matches the actual state of compliance where the coefficient is as close to +1.0 as possible, in other words, perfect agreement. So, the agreement cells are heavily weighted. We do not want to see all the cells, both agreement and disagreement cells, equally weighted. That would indicate a random response rate and a coefficient close to 0.0.

But there is another possibility which involves bias on the part of the licensing inspector in which they have certain biases or tendencies when it comes to making regulatory compliance decisions about individual rules. So, it is possible that decisions made regarding regulatory compliance could be either overall (+) positive In-Compliance or (-) negative

Not-In-Compliance when in reality the actual state of compliance is more random.

When this occurs, the coefficient falls off the range category and is not between 0 and +/-1.0 because there is no variance detected in the data. It is always biased either positively or negatively.

The UCM can be used for both reliability and validity testing as suggested in the above. Just look for different results. For validity, false positives and negatives should either be eliminated or reduced as well as possible and the remaining results should show the typical diagonal pattern as indicated by the agreement cells.

For reliability, the same pattern should be observed as in the validity testing above but there is an additional test in which bias is tested for. Bias will be ascertained if the patterns in the results indicate a horizontal or vertical pattern in the data with little or no diagonal indication. Bias can be found at the individual inspector level as well as at the standard level or the actual state of compliance.

In both reliability and validity testing, random results in which each of the cells are equally filled is not a desirable result either.

The following tables 5-10 depict the above relationships with results

highlighted in red: Table 5

Valid & Reliable Results	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Table 6

Random Results	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Table 7

Positive Bias Results Individual	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Table 8

Negative Bias Results Individual	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Table 9

Positive Bias Results Standard	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Table 10

Negative Bias Results Standard	(+) In Compliance	(-) Not In Compliance
(+) In Compliance	Agreement (++)	Disagreement (+-)
(-) Not In Compliance	Disagreement (-+)	Agreement (--)

Tables 5 – 10 demonstrate the different results based upon individual response rates when making regulatory compliance decisions about rules. Table 5 is what needs to be attained and tables 6 – 10 need to be avoided. Only in table 5 are false negatives and positives eliminated or avoided. In tables 6 – 10, false negatives and/or false positives are introduced which is not desirable when making validity or reliability decisions.

Table 6 results clearly indicate that a great deal of randomness has been introduced in the regulatory compliance decision making in which the individual licensing inspector decisions do not match reality. Tables 7 and 8, demonstrate bias in the decision-making process either positively (inspector always indicates in compliance) or negatively (inspector always indicates out of compliance). It is also possible that the standard being used has bias built into it, this is less likely but is still a possibility. The results in Tables 9 and 10 demonstrate where this could happen.

All these scenarios need to be avoided and should be monitored by agency staff to determine if there are patterns in how facilities are being monitored.

Uncertainty-Certainty Matrix for Differential Monitoring Studies

The purpose of this part of the research abstract is to explore the possibility of utilizing the Uncertainty-Certainty Matrix (UCM) not only in validation and reliability studies in licensing decision making but also with differential monitoring studies. The UCM has been proposed for use in licensing decision making but this would be an extension of this thinking to studies that involve validating licensing decisions such as when key indicators are used in comparison with comprehensive reviews of rules, and in the development of risk rules as part of the risk assessment methodology. This new Differential Monitoring 2x2 Matrix can also be used to depict the relationship between full and substantial regulatory compliance and the nature of rulemaking.

The basic premise of the DMM: Differential Monitoring Matrix is similar to the original thinking with the UCM but there are some changes in the formatting of the various cells in the matrix (see Table 11). When it comes to regulatory compliance decision making a 2 x 2 matrix can be drawn with the possible outcomes as is indicated in Table 11 where each individual rule is either in (+) or out (-) of compliance. Also, there is the introduction of a high regulatory

compliant group (+) and a low regulatory compliant group (-) which is different from the original UCM.

Table 11

DMM Matrix	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

By utilizing the format of Table 11, several key components of differential monitoring can be highlighted, such as key indicators and risk assessment rules, as well as the relationship between full and substantial regulatory compliance.

Regulatory compliance is grouped into a high group (+), generally this means that there is either full or substantial regulatory compliance with all rules. The low group (-) usually has 10 or more regulatory compliance violations. Individual rules being in (+) or out (-) of regulatory compliance is self-explanatory.

Tables 12-18 below will demonstrate the following relationships:

Table 12 depicts the key indicator relationship between individual rules and the high/low groups as indicated in red. In this table, the individual rule is in compliance with the high group and is out of compliance with the low group. This result occurs on a very general basis and should have a .50 coefficient or higher with a p value of less than .0001.

Table 13 depicts what most rules look like in the 2x2 DMM. Most rules are always in full compliance since they are standards for basic health and safety for individuals. This is especially the case with rules that have been weighted as high-risk rules. Generally, one never sees non-compliance with these rules. There will be a substantial number of false positives (+-) found with high-risk rules but that is a good thing.

Table 14 depicts what happens when full compliance is used as the only criterion for the high group. Notice that the cell right below (++) is eliminated (-+). This is highly recommended since it eliminates false negatives (-+) from occurring in the high group. As will be seen in Table 5, when substantial compliance is used as part of the high group sorting, false negatives are re-introduced. If possible, this should be avoided, however in some cases because of the regulatory compliance data distribution it is not always possible where not enough full compliant programs are present.

Table 15 depicts what occurs when substantial compliance is used as part of determining the high group. False negatives can be reintroduced into the matrix which needs to be either eliminated or reduced as best as possible. If substantial compliance needs to be used in determining the high group, then there is a mathematical adjustment that can be made which will impact the equation and essentially eliminate false negatives mathematically (see the research note at the end of this research abstract).

Table 16 depicts what happens if the individual rule is particularly difficult to comply with. Both the high performers as well as the low performers are out of compliance with the rule.

Table 17 depicts a situation where the programs are predominantly in a low group with few at full or substantial regulatory compliance which is indicative of poor performing programs. Very

honestly, this is generally not seen in the research literature, but it is a possibility and one to be in tune with.

Table 18 depicts a terrible individual rule which predicts just the opposite of what we are trying to do with programs. Obviously, this rule would need to be rewritten so that it fits with the essence of regulatory compliance in helping to protect individuals.

The following tables 12-18 will depict the above relationships with results

highlighted in red:

Table 12

Key Indicators	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Table 13

Risk Rules	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Table 14

Full Compliance	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance		(--)

Table 15

Substantial Compliance	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Table 16

Very Difficult Rule	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Table 17

Poor Performing Programs	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Table 18

Terrible Rule	High Group (+)	Low Group (-)
(+) Rule is In Compliance	(++)	(+-)
(-) Rule is Not In Compliance	(-+)	(--)

Tables 12 – 18 demonstrate the different results based on the relationship between individual regulatory compliance and if a program is either a high performer or a low performer. These tables are provided as guidance for understanding the essence of differential monitoring and regulatory compliance which has various nuances when it comes to data distributions. This research abstract hopefully can be used as a guide in determining from a data utilization point of view how to make important regulatory compliance policy decisions, such as: which rules are excellent key indicator rules, which are performing as high risk rules, importance of full compliance, what to do when substantial compliance needs to be employed, are there difficult rules to comply with, how well are our programs performing, and do we have less than optimal rules that are in need of revision.

Discussion

Over the past decade in doing research on the Regulatory Compliance Key Indicator Metric (RCKIm) it has become very clear that false negatives needed to be controlled for because of their potential to increase morbidity and mortality. When dealing with regulatory compliance and full compliance as the threshold for the high grouping variable in the 2 x 2 Regulatory Compliance Key Indicator Matrix (RCKIm) (see matrix below in Table 19), false negatives could be either eliminated or reduced to the point of no concern.

However, if substantial compliance rather than full compliance is used as the threshold for the high grouping variable in the 2 x 2 Regulatory Compliance Key Indicator Matrix (RCKIm) this becomes a problem again. There is the need to introduce a weighting factor. In utilizing the RCKIm, the following equation/algorithm is used to produce the Fiene Coefficient (FC):

$$\mathbf{FC = ((A)(D)) - ((B)(C)) / sqrt\ (WXYZ)}$$

This RCKIm needs to be revised/updated to the following to consider the need to again eliminate false negatives being generated by the results of the equation/algorithm; this can be accomplished by cubing B:

$$\mathbf{FC^* = ((A)(D)) - ((B\text{^}3)(C)) / sqrt\ (WXYZ)}$$

By this simple adjustment to cube (B = False Negatives) it will basically eliminate the use of any results in which a false negative occurs when substantial compliance is determined. The table below (Table 19) displays the variables of the Regulatory Compliance Key Indicator Matrix (RCKIm).

Table 19: RCKIm	High RC Group	RC Low Group	
KI In Compliance	A	B^3	Y
KI Violations	C	D	Z
Totals	W	X	

In the above examples, FC can be used when the High RC Group is at full regulatory compliance, but FC* needs to be used when the High RC Group is including substantial as well as full regulatory compliance. By using both equations/algorithms, it better deals with the results of the Regulatory Compliance Theory of Diminishing Returns.

The results should clearly show that only positive (+) coefficients will become Regulatory Compliance Key Indicators versus those rules that do not show any relationship to overall regulatory compliance (0), but now the negative (-) coefficients will more clearly show when any false negatives appear and clearly not include them as Regulatory Compliance Key Indicators. This is a major improvement in the Regulatory Compliance Key Indicator methodology which clearly demonstrates the differences in the results. It provides a gateway in regulatory compliance data distributions where substantial regulatory compliance is heavily present while full regulatory compliance is not. This could become a problem as the regulatory science field moves forward with the use of the Regulatory Compliance Theory of Diminishing Returns.

Conclusion

The Uncertainty-Certainty Matrix provides a useful tool for assessing the effectiveness of licensing decision making in the human services via validation and reliability studies within differential monitoring systems. It is hoped that licensing researchers and regulatory scientists will experiment with it and test it out in different arenas. It appears to have broad applicability across regulatory disciplines.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

A reprint from

American Scientist

the magazine of Sigma Xi, The Scientific Research Honor Society

This reprint is provided for personal and noncommercial use. For any other use, please send a request to Permissions, American Scientist, P.O. Box 13975, Research Triangle Park, NC, 27709, U.S.A., or by electronic mail to perms@amsci.org.
©Sigma Xi, The Scientific Research Hornor Society and other rightsholders

Finding the Rules that Work

An emerging paradigm promises to close the gap between regulatory compliance scores and the quality of childcare services.

Richard Fiene

An old fable recounts how a father and son, taking a donkey to market to sell it, encounter a string of critical villagers who each inform the pair they're "doing it wrong." Their efforts to please each subsequent critic end, absurdly and tragically, with them carrying the beast of burden themselves, ultimately causing its death.

Like the advice of those villagers, regulations are proffered in the name of safety and good practice. And, like that father and son, programs that try to follow every single rule to the letter may soon find themselves too weighed down to achieve (or perhaps even recall) what they set out to do. As the saying goes, "When you're up to your behind in alligators, it's hard to remember that you set out to drain the swamp."

In my four decades as a regulatory scientist studying childcare, I've seen this pattern play out time and again: In the lead-up to evaluations, staff at perfectly compliant programs spend so much time dotting i's and crossing t's that they have little left over for working with classrooms or teachers, whereas staff at slightly less compliant facilities, though equally careful about observing rules, fuss less with paperwork and work more with teachers on improving skills and curriculum.

Needless to say, developmentally appropriate curricula change kids' lives; boasting a perfect record does not. This observation neither dismisses the 200 to 400 rules and regulations set by respective U.S. states nor undermines the importance of complying with them, either as individual rules or in the aggregate. And full compliance does improve safety. But, as data gathered by my research team repeatedly demonstrates, a vague, uncomfortable gap separates full, costly regulatory compliance from program quality.

It is never about more or fewer rules; it is about which rules are really productive and which are not.

Moreover, early care and education providers often voice concerns that licensing inspectors inconsistently administer and apply particular rules. At issue, then, are not regulations' overall value per se, but rather the value of individual rules relative to fanatical box-checking. Given their limited resources, how can the early care and education fields get the most bang for their buck?

Such a discussion is long overdue. The unequal worth of many general licensing and quality standards, including those driven by a regulatory political bent rather than empirical evidence, produce markedly uneven developmental outcomes for kids. Today, an outcomes-based scientific reference frame is already influencing the human services industry (childcare, child welfare, and child and adult residential services), particularly in the early care and education fields (childcare centers and family childcare homes for children between infancy and 12 years old). The point of my team's approach, which I call the *theory of regulatory compliance*, is not to ask whether we need more or fewer rules, or more thorough or less thorough compliance, but rather to evaluate which rules truly prove effective.

Modernizing Measurement

Regulatory scientists use tools, standards, and methodologies to assess the safety, efficacy, and quality of programs under government regulation. Ideally, they help regulatory agencies achieve the best possible public health and safety outcomes.

The regulatory science field has a lot of ground to make up. At about 30 years old, it lags its subject matter by a good century (Pennsylvania passed the first orphanage licensing law in the United States almost 140 years ago). Human services licensing grew slowly prior to the late 1960s to early 1970s, when American President Lyndon B. Johnson began the Great Society initiatives such as Head Start, which kicked off the rapid multipli-

QUICK TAKE

Contrary to historical assumptions, the quality of childcare programs does not increase linearly as their compliance with rules and regulations approaches 100 percent.

All-or-nothing, one-size-fits-all approaches to compliance and licensing generate skewed data, raise risks of false negatives and false positives, and burden staff with bureaucratic tasks.

Substantial regulatory compliance is an alternative approach that emphasizes compliance with the most productive rules, preserves safety, and allows staff to concentrate more on children.

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.

DGLimages/Shutterstock

Staff of fully compliant childcare programs say they spend too much time box-checking and not enough working with teachers, whereas staff at slightly less compliant facilities, though equally scrupulous, bother less with form-filling and spend more time in the classroom. An outcomes-based substantial regulatory compliance approach lets licensors strike that balance.

cation of childcare programs. Those decades also saw human services, especially childcare, begin transforming from cottage industries, with program monitoring and measurement conducted qualitatively via case notes and anecdotal records, to more rigid systems that entailed oversight, case reviews, and state agency inspections. In the 1970s, these systems, which often varied from state to state, gave way to improvements brought by the Federal Interagency Day Care Requirements.

The watershed moment for regulatory science as it pertains to children's programs came in the 1980s. The previous decade's major childcare expansion in the United States had created a backlog of licensing assessments, caused unmanageable monitoring delays, and laid bare the logistical limits of case studies. These factors, combined with advances in computing, led states to introduce an empirical, quantitative, and instrument-based approach, complete with sophisticated software systems designed by state agencies and private vendors to track regulatory compliance and quality assessment data. Empirical evidence not only moved regulatory science from qualitative to quantitative analysis, it also revealed surprising patterns.

But first, some background: As the U.S. Department of Health, Education, and Welfare took over running the show for all U.S. early care and education programs in the 1970s, *uniform program monitoring* had become the rule. Uniform monitoring derived from the philosophical assumption that fuller regulatory compliance would produce, linearly, better quality across U.S. early care and education programs. As the former went up, so would the latter. From a public policy standpoint, this notion sounds aspirational, but sensible: Any licensing agency looks for service quality to increase as its rules, regulations, and standards are followed.

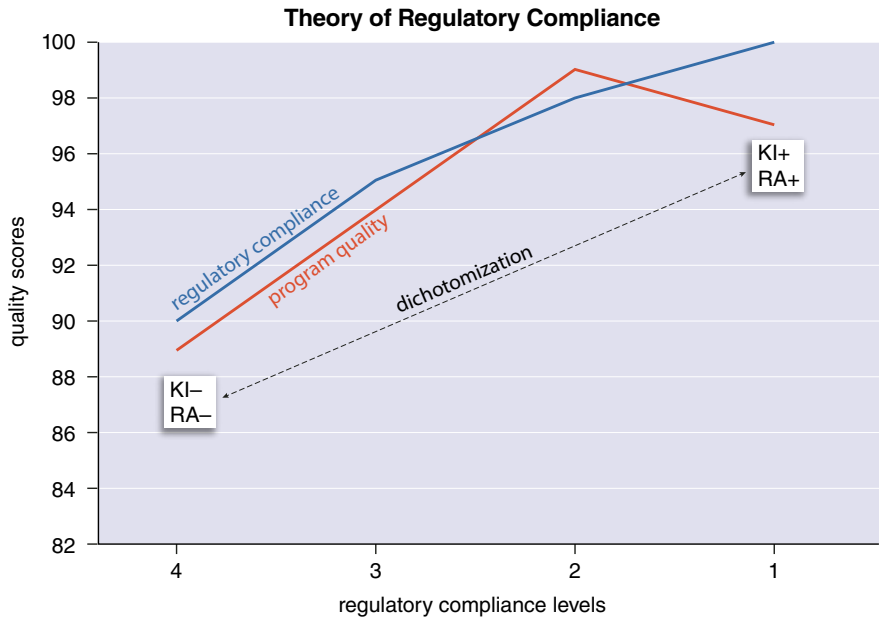
But as expert opinion and anecdotal evidence gave way to better-designed studies and empirical data, and as larger studies became possible thanks to data computerization by state licensing agencies, cracks appeared. When researchers compared violations found during licensing reviews and inspections to the quality of the violating programs, they found that a linear relationship did indeed exist between quality and compliance—but only as one moved from low compliance levels to *substantial regulatory compliance* (that is, 98–99 percent). Between that and 100 percent compliance, quality consistently plateaued and, as some 2010s replication studies suggested, even showed diminishing returns.

A New Paradigm

These results called into question the notion that state agencies should issue licenses solely to fully compliant programs. If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly, we needed to rethink our program evaluation strategies.

In the United States, state licensing and regulatory agencies establish childcare regulations, but federal agencies such as the Office of Child Care and the Administration for Children

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.



adapted from Richard Fiene

This graph shows the quality scores (*y*-axis) associated with four categories of regulatory compliance (*x*-axis, defined by the number of rules violations, ranging from 0 [Level 1] to 10 or more [Level 4]). Note that compliance scores (*blue line*) and quality scores (*red line*) rise together, but only until substantial compliance (99-97 percent compliance with all rules (Level 2) is reached. This finding argues for the adoption of substantial compliance as a standard, and for utilizing differential/relative monitoring to better capture nuances of quality and more efficiently allocate resources. The alternative—a punitive, gatekeeping licensing approach requiring full compliance (a yes/no proposition)—has led to highly skewed data. Here, the author has split (dichotomized) these skewed data into two extremes: Programs with regulatory compliance scores in the top 5-10 percent (*upper right*, labeled KI+/RA+ to indicate positive key indicator and risk assessment findings) and the bottom 5–10 percent (*lower left*, labeled KI-/RA-). The graph shows how scores in key indicators and risk assessment effectively predict program quality.

and Families also influence rules, as does Congress through its funding purse strings. Sometimes cities and counties, too, set regulations or standards, especially concerning physical environment, health, safety, and zoning. (Here, the term "regulations" means those defined by the National Association for Regulatory Administration's Licensing Curriculum.)

For an individual program or facility to operate, a state licensing agency must judge that it follows these standards. Examples include certifications for teacher qualifications, first aid, CPR, and the facility environment, along with requirements for ongoing training and professional development. State licensing staff evaluate compliance via inspections, document reviews, audits, and interviews, usually on a yearly basis. Inspections check for health, safety, cleanliness, educational standards, and staff-to-child ratios, as well as less obvious standards such as playground and transportation safety. Noncompliant programs may face fines, mandated corrective actions, training, or technical assistance, or may undergo license suspension or even permanent closure.

Licensing requirements vary depending on the childcare offered (such as family childcare homes, center-based care, or school-based programs), with larger centers typically facing more stringent requirements. Along with compliance ratings and violations issued by licensing inspectors, these facilities voluntarily seek ratings from quality initiative offices within human services agencies.

Here, and in my research, I primarily deal with center-based care programs, but the findings apply to other service types as well, such as family childcare homes and school-age programs, as well as human services categories such as child residential, child foster care, adult residential, and adult personal care homes. My data and research concern the relationship between quality and compliance, and how to improve it. They stem from studies of hundreds of programs I conducted at the state level from the 1970s through the 2010s, when I directed various research and training institutes at Pennsylvania State University. In these controlled and replicated studies, trained observers collected both regulatory data and program quality data from eight states, three Canadian provinces, and the U.S. Head Start program. The work ran the gamut, from site selection via stratified random samples, to dispatching data collectors to specific programs, to providing individual states with an overall blueprint describing how to conduct their studies.

Initially, the ceiling effect between regulatory compliance and program quality came as a surprise; we did not predict that full compliance would fail to outperform substantial compliance. It also drew pushback from the licensing field. Thus, I replicated the study many times over to assess my assumptions. But the finding persisted: Program quality scores rise with regulatory compliance until programs reach substantial compliance, after which quality declines. Although until 1980 states required childcare programs to show full compliance and zero violations, since 2015 most states have allowed licensing for facilities that are substantially compliant.

Differential Monitoring

If substantial compliance with some rules rather than full compliance with all rules best ensures the childcare program quality, then the question naturally arises: "Which rules?" Conceivably, some rules should weigh more heavily than others—say, the ones that data show most closely relate to safety and quality. Such is precisely the idea behind *differential monitoring*.

Differential monitoring emerged in 1979 during my discussions with federal agencies such as the Administration for Children, Youth and Families and the Children's Bureau, who felt dissatisfied with the traditional uniform monitoring approach. They knew about my team's work in Pennsylvania and invited me to give a series of talks to their staff. The result was a move away from the older, one-size-fits-all approach to differential methods focused on *key indicators* and *risk assessments*.

Key indicators are statistical predictors of overall compliance—rules that, if a facility follows them, strongly suggest they will follow other rules as

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.

well. They very efficiently determine a facility's overall regulatory compliance without requiring a comprehensive inspection. Far from negligent, this approach works because not all rules are created and monitored equally.

Risk assessment focuses on those rules and regulations which, when breached, place children at greatest risk, such as rules that deal with supervision or hazardous materials handling, among others. Generally, jurisdictions, states, and provinces engage major early care and education stakeholders (service providers, parents, advocates, and licensing staff) in weighting rules or regulations based on their risks to children's health and safety. Commonly, participants assign weights via a *Likert scale*—a common survey and questionnaire tool that lets respondents indicate the strength of their agreement or disagreement (or, in this case, their assessment of risk) with a statement about attitudes, opinions, or perceptions. The weights range from 1 to 10, where 1 indicates little risk if a program fails to follow the specific rule or regulation and 10 corresponds to high risk. Rules heavily weighted as associated with sickness, injury, or death join the risk assessment rules measured by inspectors in every differential monitoring review.

As an aside, I should point out that full compliance remains the standard for maintaining health and safety. So why incorporate risk assessments into differential monitoring and, by extension, the substantial compliance paradigm, as its own separate metric? In truth, I had no such intention when I wrote my 1985 research papers about differential monitoring and the theory of regulatory compliance. Rather, risk assessment morphed from a way to provide the needed data variance for key indicator scoring into its own submethodology. As it found its way into the implementation of national standards and guidelines, risk assessment subsequently emerged as a separate methodology.

Our findings repeatedly show that using the combined methodologies of key indicator predictor rules and risk assessment rules to identify the "right rules" and to ensure compliance with them, rather than to seek full compliance, makes the differential monitoring approach the most effective and efficient program monitoring system. Also, studies show that abbreviated, targeted, and focused reviews take approximately 50 percent less time than comprehensive reviews.

Compliance Measurement Systems

scoring level	individual rule		aggregate rules	individual rule
scale	instrument based	*scale*	differential	integrated
7	full compliance	7	full compliance	exceeds compliance
–	–	5	substantial	full compliance
–	–	3	mediocre	substantial
1	out of compliance	1	low	mediocre/low

adapted from Richard Fiene

This table compares different approaches to measuring compliance: A licensing-focused approach in which programs are classified as either compliant or noncompliant based on rules violation counts, with no middle ground (*columns 1 and 2*), and a more nuanced ordinal approach using a Likert scale. This experimental metric, called the Regulatory Compliance Scale (*column 3*), is currently being tested at the aggregate rule level (*column 4*) and may be expanded to the level of individual rules (*column 5*) in the future. Note that aggregate rule scores are not equal to the sum of all individual rule scores because not all rules are created or administered equally.

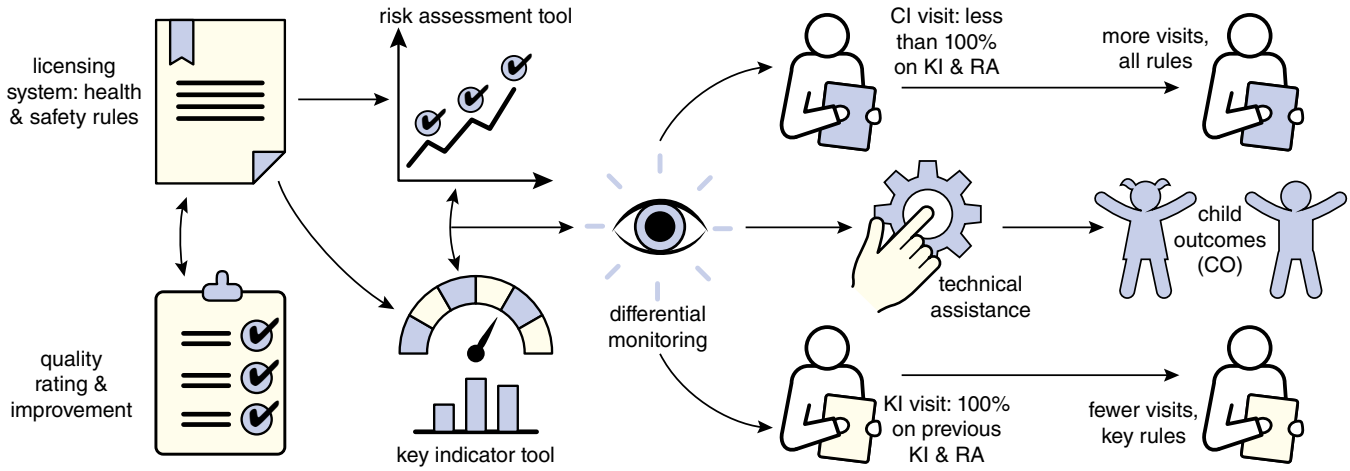
Unfortunately, although many licensing bodies use risk assessment or key indicator methodologies, few use both. *Monitoring Practices Used in Child Care and Early Education Licensing*, a federal accounting of how states conduct program monitoring, reported that 10 states used key indicators, 17 states used risk assessments, and only one state used both. Hopefully, this pattern will change as the regulatory science field matures over the coming decades.

If, as data suggested, substantially compliant programs provided the same or better care as fully compliant ones, then clearly we needed to rethink our program evaluation strategies.

Since I first proposed it in the mid-1980s, the theory of regulatory compliance has faced numerous critics in the human services licensing field, especially among advocates of uniform monitoring and full compliance. Only after years of licensing validation studies conducted by my team and others repeatedly demonstrated that full compliance did not produce the highest quality did states begin licensing programs in substantial rather than full regulatory compliance. Today, although various U.S. states apply the differential monitoring review approach unevenly, nearly all have adopted the policy of granting licenses for substantial rather than full compliance. The latest revision of the legislation for the Child Care and Development Block Grant (a U.S. federal funding program that helps states, territories, and tribes assist low-income families access affordable childcare) cites differential monitoring as an alternative to uniform program monitoring.

Of all the approaches and methodologies that flow from the theory of regulatory compliance, differential monitoring most significantly alters the program monitoring, inspection, review, and licensing landscape. Its reviews occur just as often as do uniform monitoring assessments but focus specifically on rule breaches shown to place children at risk. That said, differential monitoring did not replace but rather supplemented its predecessor: Comprehensive reviews must still occur every three to four years to validate the performance of key indicators and risk assessment rules. But what does that report card look like in terms of analyzable data?

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.


licensing system: health & safety rules
risk assessment tool
quality rating & improvement
key indicator tool
differential monitoring
CI visit: less than 100% on KI & RA
more visits, all rules
technical assistance
child outcomes (CO)
KI visit: 100% on previous KI & RA
fewer visits, key rules

Barbara Aulicino

This illustration shows the various components that contribute to a differential monitoring approach and how agencies can use them to evaluate the effectiveness and validity of different approaches. Differential monitoring allocates resources based on risk assessment (client morbidity and/or mortality) and key indicators (rules whose compliance is strongly predictive of program quality). These data, provided by mandatory licensing processes and voluntary quality rating services, reveals which programs are highly compliant with key rules (though not all rules) and therefore require fewer visits versus programs that are less compliant and require additional visits and technical assistance to achieve similar child outcomes.

Rethinking Nominal Data

Traditionally, licensing data are categorical (sorted into groups such as "approved" or "denied"), unordered (there's no built-in way for such groups to be sequenced), and mutually exclusive (state agencies cannot simultaneously deem a facility both "approved" and "denied"). In statistical terms, such data are nominal, like a table listing cars by make or model; you cannot "do math" on such a table like you can on, say, on a table listing automobile curb weights and fuel economies. It is also binary: A program either follows a rule, or it doesn't.

Presently most jurisdictions deal in these absolutes and exclude gray areas. This approach, much like uniform program monitoring and full compliance, makes intuitive sense: We create rules and regulations because we believe in the value of following them, and because licenses mean nothing if licensees are not held to a standard. But here again, we must look deeper and ask, "What consequences follow from this either/or approach to measuring compliance, and who decides whether or not a particular box gets checked?"

Let's begin with the latter question. In an ideal world, judgments made by assessors would perfectly reflect a program's actual regulatory compliance state. But research that tests reliability and replicability in the licensing field empirically shows a concerning degree of disagreement when a second observer validates the decision regarding regulatory compliance. These disagreements suggest a worrying number of false positives and false negatives.

A false positive occurs when a program follows a rule or regulation, but the assessor rules that the facility is noncompliant (which might sound backwards, but the metric is *non*compliance, not compliance, so finding a false violation means finding a false positive). But even more concerning are false negatives, in which an evaluator says a program complies with a rule that it breaches, thereby placing clients at risk. Detecting false negatives is one of the chief reasons we periodically validate the predictive value of key indicator rules through comprehensive reviews.

As for the first question, the answer is simple: Nominal, binary licensing data is severely skewed. Upon reflection, the reason becomes obvious. When a regulated industry such as childcare mandates compliance before a program can operate and excludes gray areas, most facilities will achieve full compliance or lose their licenses. Because unlicensed providers don't last long, the childcare sector produces data that skew toward licensed programs. To grasp such skewed continuous or multicategory data, we must first dichotomize it into two distinct groups.

Such sorting into piles raises statisticians' hackles; unless carefully done, it accentuates differences and forces tradeoffs between precision and sensitivity, which can mean swapping false positives for false negatives. But the nature of licensing data—a skewed collection of mostly or fully compliant programs dumped in a single bucket—makes the split both necessary and warranted. By setting a threshold of certainty or agreement among evaluators, we can more effectively reduce false negatives, that is, cases in which evaluators say a program follows a rule when it doesn't.

This need becomes even clearer when one considers the demands posed by differential monitoring and its methodologies, key indicators, and risk assessments. For a program to receive licensure, it is not enough to ask if it "complies enough overall"; we must also know if it follows the specific rules that most ensure safety. By comparing highly compliant programs only with low-compliant programs, we accentuate the differences between the two and bolster our data analyses as well as overall safety. This comports well with licensing decision-making, which can consider a program compliant or noncompliant not only in aggregate, but with respect to *individual* rules.

Infusing Quality

The all-or-nothing approach to regulatory compliance and licensing fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from a false assumption that program quality increases in step with 100 percent compliance. But I am far from the first

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.

to notice that approach's weaknesses in evaluating how good a program or facility actually is. Indeed, its shortcomings helped drive the creation of a separate industry of voluntary accreditation programs such as the National Association for the Education of Young Children, state-run quality rating and improvement systems, and third-party tools and assessments. It's time we folded quality assessments into regulatory compliance.

I have already explained how the theory of regulatory compliance improves program quality and safety by focusing on substantial, not full, compliance and by using differential monitoring to ensure programs follow the most protective and impactful rules. But to further cast off the limitations and lopsidedness of a uniform monitoring and full compliance mindset, and to make room for data capable of tracking quality, we must also replace rigid either/or logic with a more nuanced ordinal measurement: a scaling technique.

Recall that assessors can evaluate compliance in two ways: They can consider aggregate rules—collections of rules that fall into categories such as staffing or safety practices—or individual rules. Each has its own studies and research literature. Research on aggregate rules from the 1970s, 1980s, and the 2010s established substantial compliance as a "sweet spot" of best outcomes and showed that the time had come to replace nominal metrics (such as "compliant" and "noncompliant") with ordinal ones (such as "98 percent compliant").

Inspired by this research, I have proposed replacing older nominal techniques with an ordinal scale like the Likert scale already used in quality measurements (usually but not always ranging from 1–7, with 1 being inadequate and 7 being excellent). This technique, currently under review by the National Association for Regulatory Administration, will help reviewers consider the importance of substantial compliance. Moreover, it will add the currently absent quality elements to each rule and regulation. However, this approach involves aggregate rules only; further research is needed to determine if the same shift from nominal to ordinal metrics should also occur at the individual rule level.

Should those findings bear out the value of evaluating individual rules via the 1–7 regulatory compliance scale, I propose that it should contain the following categories: exceeding full compliance, full compliance, substantial compliance, and mediocre compliance (*see figure on page 19*). These categories differ from the aggregate rule compliance scale currently under evaluation (full, substantial, mediocre, and low compliance) because aggregate compliance only considers health and safety elements, whereas an individual scale would also take quality into account.

Research supports the value of transitioning from uniform monitoring and full compliance to differential monitoring and substantial compliance. Practice has shown the value of retaining the older to help ensure the validity of the newer. Looking to the future, I believe we can further improve compliance evaluations by developing and evaluating *integrative monitoring*, which incorporates program quality into rule formulation and moves the key indicators from predicting compliance to forecasting quality.

The all-or-nothing approach fails as a standard because it generates skewed data, raises the risks of false negatives and false positives, and springs from the false assumption that program quality increases in step with 100 percent compliance.

Looking Forward

The regulatory compliance scale is a new and evolving metric. It transforms licensing data from a mere violation tally into a more useful and intuitive scale, one more consistent with the program quality measurements supported by research. Hereafter, I hope that the approach will incorporate quality measurements and more nuanced weighting into the evaluation of individual rule compliance. But discussions are just beginning, and this shift will pose a substantial challenge for agencies, which must also cope with the aftermath of the COVID-19 pandemic and a rising tendency toward deregulation.

The theory of regulatory compliance concerns the relationship between regulatory compliance and program quality, not health and safety, where full compliance remains the goal. It is, however, the preferred methodology for eliminating false negatives and decreasing false positives. Add to that the fact that the theory of regulatory compliance predicts a nonlinear relationship between compliance and quality but a linear relationship linking regulatory compliance and safety, and regulatory scientists clearly have our work cut out for us. Untying this knot will require greater collaboration between the historically siloed public policy worlds of licensing, accreditation, quality rating and improvement systems, and professional development systems.

I hope that the regulatory science field takes these paradigm shifts into consideration as it builds licensing decision-making systems and considers how states issue licenses. And although this work deals primarily with my own experience in the early care and education field, I wonder if other human service sectors, such as the foster care or child and adult residential areas, demonstrate similar patterns. Other disciplines that deal with regulations and compliance may similarly find it fruitful to discuss the nuances of their own evaluation metrics in order to achieve the best overall outcome with the most efficient use of limited resources.

Bibliography

Fiene, R. 2023. Licensing measurement and monitoring. National Association for Regulatory Administration licensing curriculum. Published online www.naralicensing.org/key-indicators

Fiene, R. 2022. Regulatory compliance monitoring paradigms and the relationship of regulatory compliance/licensing with program quality: A policy commentary. *Journal of Regulatory Science* 10:1–7.

Fiene, R. 2019. A treatise on regulatory compliance. *Journal of Regulatory Science* 7:1–3.

Richard Fiene is an emeritus professor of psychology at the Pennsylvania State University and has had a lengthy career in regulatory science and developmental psychology, in which his research has led to the development of national early care and education standards. He directs the Research Institute for Key Indicators Data Lab. Email: RFiene@RIKInstitute.com

© 2025 Sigma Xi, The Scientific Research Honor Society. Reproduction with permission only. Contact perms@amsci.org.

TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http:\\journalofregulatoryscience.org

Regulatory Science

Journal of Regulatory Science 7 (2019) 1–3

A Treatise on the Theory of Regulatory Compliance

Richard J. Fiene[a,*]

[a]*Edna Bennett Pierce Prevention Research Center, Pennsylvania State University, 305 Templar Drive, Elizabethtown, Pennsylvania 17022*

Abstract

This treatise provides some insights into certain assumptions related to regulatory compliance and the implications for regulatory researchers and policy-makers for the future development of rules and regulations. Once regulatory compliance decision making moves from requiring full compliance with all rules to a substantial regulatory compliance decision making approach, the measurement and monitoring systems employed to assess programs and facilities change dramatically.

Keywords: regulatory compliance, risk assessment, key indicators, licensing, monitoring, measurement

1. Introduction

Regulatory compliance is a sub-discipline within regulatory science that focuses on measurement, monitoring systems, risk assessment, and decision making based on regulatory compliance scoring. Regulatory compliance is dominated by nominal scale measurement, that is, either a facility is in or out of compliance with specific rules. There is no middle ground with regulatory compliance as there is with most quality measurements, which are generally made on an ordinal scale. However, some regulators feel that certain regulations are not or should not be subjected to nominal measurement.

A factor with regulatory compliance data is that they generally follow a very skewed frequency distribution, which limits analyses to non-parametric statistics. Because of the skewed data distribution, dichotomization of data is warranted, given the lack of variance in the regulatory compliance frequency distribution - the majority of facilities [1] are either in full or substantial regulatory compliance.

An assumption within regulatory compliance is that full regulatory compliance, that is, 100 percent compliance with all rules [2], is the best (i.e., risk is minimized) possible scenario for the services being delivered and assessed. It is also assumed that all promulgated rules have an equal weight in their relative impact on the desired service delivery model, although this thinking has been changing over time regarding how rules are reviewed and complied with. This short treatise will examine the past 40 years of research delving into regulatory compliance measurement, and will provide some guidance to regulatory researchers and policy-makers as they move forward with both research and policy development related to rules. The data from these research studies have led to a Theory of Regulatory Compliance that demonstrates that substantial regulatory compliance - and not full regulatory compliance - is a more effective and efficient public policy as it relates to decision making on monitoring and licensing.

The results reported herein are drawn from human services delivery systems in the United States and Canada, such as early care and education, as well as child and adult residential services. The results are from state and provincial level licensing systems involving over 10,000 facilities serving over 100,000 clients. All the data are part of an international regulatory compliance database (https://data.mendeley.com/datasets/kzk6xssx4d/1) maintained at the Research Institute for Key Indicators and the Pennsylvania State University.

2. Methods

Alternate methodologies, logic models, and algorithms were developed directly from the Theory of Regulatory Compliance once it was determined that substantial regulatory compliance produced better results than full regulatory compliance. These methodologies created a differential monitoring or targeted monitoring approach based on risk assessment, which measures client morbidity and/or mortality when individual rule

*Corresponding author: Richard J. Fiene, Email: rjf8@psu.edu, Phone: 717-598-8908, ORCID iD: http://ORCID: 0000-0001-6095-5085.

[1] The term "facilities" is used when referring to programs and/or facilities.

[2] The term "rules" is used when referring to rules and/or regulations.

non-compliance is assessed, and the determination of key statistical predictors for overall regulatory compliance [3].

Briefly, the above methodologies provide cost-effective and efficient means for the ongoing monitoring of human service delivery systems by selecting and reviewing only those rules that either have a positive impact on clients, statistically predict overall regulatory compliance, or protect the health and safety of clients [3]. Based on regulatory compliance historical data, decisions could be made as to the frequency and depth of the reviews or inspections. Abbreviated reviews (inspections in which a subset of rules are measured), such as licensing key indicator rules or risk assessment rules, would only be done in those facilities having a history of high regulatory compliance. Those facilities with a history of high regulatory non-compliance would continue to receive full regulatory compliance reviews as they did in the past.

3. Results

Prior to 1979, it was always assumed that there was a linear relationship between regulatory compliance measures and program quality measures of human service facilities. In a study conducted in that year, which compared results from early care and education programs, in particular child care centers, this assumption did hold up when one went from low regulatory compliance to substantial regulatory compliance. However, the results from substantial regulatory compliance to full (100 percent) regulatory compliance did not show the same linear relationship. Rather, it showed that those programs that were in substantial instead of full compliance were actually scoring higher on the program quality measures.

Since 1979, this result has been replicated in many other early care and education delivery system studies, both nationally in the United States (Head Start) [1] and in several states (Georgia, Indiana, Pennsylvania) [2]. In all these studies, one finds a non-linear - rather than a linear - relationship between regulatory compliance and the overall quality of the facilities being assessed.

4. Discussion

Based on the results above, there are several assumptions within regulatory compliance that need to be reconsidered:

1. Public policies that require full (100 percent) compliance with all rules may not be in the best interest of the clients being served, nor an effective use of limited regulatory resources. Potentially, emphasis on substantial regulatory compliance may be a more effective and efficient public policy related to client outcomes when it comes to their health, safety, and quality of life. Note that substantial compliance is still very high regulatory compliance (99-97 percent compliance with all rules) and produces positive client outcomes. As stated above, regulatory compliance data are extremely skewed and not normally distributed. There is very little variance in the data and the majority of programs are in either full or substantial regulatory compliance.

2. If a jurisdiction focuses on a substantial regulatory compliance public policy it opens up many system enhancements, such as differential or targeted monitoring, risk assessment analysis, and statistical key indicator rules that have been demonstrated to be cost effective and efficient approaches to reviewing program performance. In a full regulatory compliance public policy approach, none of these system enhancements can be employed, with the possible exception of the key indicator approach as delineated in number four below.

3. If a jurisdiction takes the position that all rules are not equal, then a risk assessment or weighting approach becomes an alternative based on the assumption that certain rules place clients at greater risk of death, serious injury, or other types of harm.

4. Even if a jurisdiction does not have a licensing law that allows issuing licenses on the basis of substantial compliance, there is the possibility that key indicators could still be used for abbreviated reviews or inspections, if there is no prohibition in statute or regulation that expressly forbids the use of this approach, since key indicators statistically predict full regulatory compliance. In other words, all rules are statistically predicted to be in regulatory compliance based on the results of the key indicators. Therefore, technically, all rules have been reviewed albeit short of a full review or inspection.

5. Based on previous research, utilizing a risk assessment approach along with a key indicator approach is the most cost effective and efficient differential monitoring system model. The reason is that both predictive rules and those rules that place clients at greatest risk are always assessed when a site visit review or inspection is done. Many more jurisdictions use a risk assessment approach at this point, but there is a loss of predictive regulatory compliance by just using it.

6. Based on previous regulatory compliance history, only those facilities in high regulatory compliance would be eligible for abbreviated key indicator and risk assessment reviews, whereas those with a history of high regulatory non-compliance would continue to receive full regulatory compliance reviews. This gets at the essence of the differential monitoring approach, which is cost neutral. Regulatory resources may then be re-allocated from the abbreviated reviews to more in-depth full regulatory compliance reviews.

7. Based on the use of the key indicator and risk assessment methodologies within a differential monitoring approach, it is possible to identify over multiple jurisdictions if there are generic rules that meet the criteria of risk abatement and prediction. Such an application has occurred in the United States with the creation of early care

and education standards entitled *Caring for Our Children Basics*, published by the Administration for Children and Families, US Department of Health and Human Services (2015).

5. Conclusion

Regulatory compliance is relatively new in applying empirical evidence and basic scientific principles to its decision making. In the past, it had been dominated by case studies and long narrative reports that did not lend themselves to quantitative analysis. There is a need to more clearly apply empirical evidence and the scientific method to rule development. Certain assumptions, such as full regulatory compliance as a sound public policy, are lacking in empirical evidence. This treatise on a theory of regulatory compliance is provided for its heuristic value for both regulatory researchers and policymakers in rethinking some basic regulatory compliance assumptions. It is not about more or less, rules but finding the "right rules" that protect clients, predict overall regulatory compliance, and produce positive client outcomes.

6. Declaration of Conflicting Interest

The authors declare no conflicts of interest.

7. Article Information

This article was received March 14, 2019, in revised form April 11, 2019, and made available online May 9, 2019.

8. References

[1] Fiene, R. (2013). *Office of Head Start Key Indicator Project Report.* Retrieved from https://drfiene.files.wordpress.com/2018/04/rheadstart-report.pdf

[2] Fiene, R. (2014). *Georgia Child Care Licensing Study: Validating the Core Rule Differential Monitoring System.* Retrieved from https://drfiene.files.wordpress.com/2018/04/2014-georgia.pdf

[3] Fiene, R. (2016). *Early Childhood Program Quality Indicator and Improvement Model (ECPQIM) and Differential Monitoring Logic Model and Algorithm (DMLMA) Readings*. Retrieved from https://drfiene.files.wordpress.com/2018/03/1ecpqims-book-of-readings.pdf

Regulatory Compliance Monitoring Paradigms and the Relationship of Regulatory Compliance/Licensing with Program Quality: A Policy Commentary

Richard Fiene, PhD [a,*]

[a] *Edna Bennett Pierce Prevention Research Center, The Pennsylvania State University*

Abstract

This policy commentary deals with two key issues within regulatory science related to the best methods for measuring regulatory compliance: Program monitoring paradigms and the relationship of regulatory compliance/licensing with program quality. Examples from program monitoring paradigms include: 1) Substantial versus Monolithic. 2) Differential Monitoring versus One size fits all monitoring. 3) "Not all standards are created equal" versus "All standards are created equal". 4) "Do things well" versus "Do no harm". 5) Strength based versus Deficit based. 6) Formative versus Summative. 7) Program Quality versus Program Compliance. 8) 100-0 scoring versus 100 or 0 scoring. 9) QRIS versus Licensing. 10) Non-Linear versus Linear. Examples from the relationship of regulatory compliance/licensing with program quality include: 1) "Do no harm" versus "Do good". 2) Closed system versus Open system. 3) Rules versus Indicators. 4) Nominal versus Ordinal measurement. 5) Full versus Partial compliance. 6) Ceiling effect versus No Ceiling effect. 7) Gatekeeper versus Enabler. 8) Risk versus Performance.

Keywords: regulatory compliance, program monitoring, licensing, program quality.

Introduction

This commentary on policy will deal with two key issues within regulatory science that need to be dealt with by licensing researchers and regulatory scientists as they think through the best methods for measuring regulatory compliance: 1) Program monitoring paradigms; 2) Relationship of regulatory compliance/licensing and program quality. The examples drawn are from early childcare and education but the key elements and implications can be applied to any field of study related to regulatory science that involves rules/regulations/standards. For the purposes of this manuscript "rules" will be used to

* Corresponding author: Richard Fiene; Email: rjf8@psu.edu;
Phone: 717-598-8908; ORCID: http://ORCID: 0000-0001- 6095-5085.

describe or refer to "rules/regulations/standards".

Program Monitoring Paradigms:

This section provides some key elements to two potential regulatory compliance monitoring paradigms (Differential/Relative versus Absolute/Full) for regulatory science based upon the Regulatory Compliance Theory of Diminishing Returns (Fiene, 2019).

As one will see, there is a need within regulatory science to get at the key measurement issues and essence of what is meant by regulatory compliance. There are some general principles that need to be dealt with such as the differences between individual rules and rules in the aggregate. Rules in the aggregate are not equal to the sum of all rules because all rules are not created nor administered equally. And all rules are to be adhered to, but there are certain rules that are more important than others and need to be adhered to all the time. Less important rules can be in substantial compliance most of the time but important rules must be in full compliance all of the time (Fiene, 2019).

Rules are everywhere. They are part of the human services landscape, economics, banking, sports, religion, transportation, housing, etc... Wherever one looks we are governed by rules in one form or another. ***The key is determining an effective and efficient modality for negotiating the path of least resistance in complying with a given set of rules***[2]. It is never about more or less rules, it is about which rules are really productive and which are not. Too many rules stifle creativity, but too few rules lead to chaos. ***Determining the balance of rules is the goal and solution of any regulatory science paradigm.***

Differential/Relative versus Absolute/Full Regulatory Compliance Paradigms: this is an important key organizational element in how rules are viewed when it comes to compliance. For example, in an absolute/full approach to regulatory compliance either a rule is in full compliance or not in full compliance. There is no middle ground. It is black or white, no shades of gray as are the cases in a differential/relative paradigm. It is 100% or zero. In defining and viewing these two paradigms, this dichotomy is the organizational key element for this paper. In a differential/relative regulatory compliance paradigm full compliance is not required and emphasis on substantial regulatory compliance becomes the norm.

Based upon this distinction between differential/relative and absolute/full regulatory compliance paradigms, what are some of the implications in utilizing these two respective approaches. Listed below are the basic implications that occur when selecting either of the two approaches on program monitoring systems: differential/relative versus absolute/full regulatory compliance paradigms.

There are ten basic implications that will be addressed: 1) Substantial versus Monolithic. 2) Differential Monitoring versus One size fits all monitoring. 3) "Not all standards are created equal" versus "All standards are created equal". 4) "Do things well" versus "Do no harm". 5) Strength based versus Deficit based. 6) Formative versus Summative. 7) Program Quality versus Program Compliance. 8) 100-0 scoring versus 100 or 0 scoring. 9) QRIS versus Licensing. 10) Non-Linear versus Linear.

1) Substantial versus Monolithic: in monolithic regulatory compliance monitoring systems, it is one size fits all, everyone gets the same type of review (this is addressed in the next key element below) and is more typical of an absolute paradigm orientation. In a substantial regulatory compliance monitoring system, programs are monitored on the basis of their past compliance history and this is more typical of a relative paradigm orientation. Those with high compliance may have fewer and more abbreviated visits/reviews while those with low compliance have more comprehensive visits/reviews.

2) Differential Monitoring versus One Size Fits All Monitoring: how does this actually look in a program monitoring system. In differential monitoring (Differential/Relative Paradigm), more targeted or focused visits are utilized spending more time and resources with those problem programs and less time and resources with those programs that are exceptional. In the One Size Fits All Monitoring (Absolute/Full Paradigm), all programs get the same type/level of review/visit regardless of past performance.

3) “Not all standards are created equal” versus “All standards are created equal”: when looking at standards/rules/regulations it is clear that certain ones have more of an impact on outcomes than others. For example, not having a form signed versus having proper supervision of clients demonstrates this difference. It could be argued that supervision is much more important to the health and safety of clients than if a form isn’t signed by a loved one. In a differential/relative paradigm, all standards are not created nor administered equally; while in an absolute/full paradigm of regulatory compliance, the standards are considered created equally and administered equally.

4) “Do things well” versus “Do no harm” (this element is dealt with in the second component to this paper below as well): “doing things well” (Differential/Relative Paradigm) focuses on quality of services rather than “doing no harm” (Absolute/Full Paradigm) which focuses on protecting health and safety. Both are important in any regulatory compliance monitoring system but a balance between the two needs to be found. Erring on one side of the equation or the other is not in the best interest of client outcomes. "Doing no harm" focus is on the "least common denominator" – the design and implementation of a monitoring system from the perspective of focusing on only 5% of the non-optimal programs ("doing no harm") rather than the 95% of the programs that are "doing things well".

5) Strength based versus Deficit based: in a strength-based monitoring system, one looks at the glass as “half full” rather than as “half empty” (deficit-based monitoring system). Emphasis is on what the programs are doing correctly rather than their non-compliance with standards. A strength-based system is non-punitive and is not interested in catching programs not doing well. It is about exemplars, about excellent models where everyone is brought up to a new higher level of quality care.

6) Formative versus Summative: differential/relative regulatory compliance monitoring systems are formative in nature where there is an emphasis on constant quality improvement and getting better. In absolute/full regulatory compliance monitoring systems, the emphasis is on being the gate-keeper (more about the gate-keeper function in

the next section on regulatory compliance/licensing and program quality) and making sure that decisions can be made to either grant or deny a license to operate. It is about keeping non-optimal programs from operating.

7) Program Quality versus Program Compliance: (this element is dealt with in greater detail in the second component of this manuscript) differential/relative regulatory compliance monitoring systems focus is on program quality and quality improvement while in absolute/full regulatory compliance monitoring systems the focus in on program compliance with rules/regulations with the emphasis on full, 100% compliance.

8) "100 – 0 scoring" versus "100 or 0 scoring": in a differential/relative regulatory compliance monitoring system, a 100 through zero (0) scoring can be used where there are gradients in the scoring, such as partial compliance scores. In an absolute/full regulatory compliance monitoring system, a 100% or zero (0) scoring is used demonstrating that either the standard/rule/regulation is fully complied with or not complied with at all (the differences between nominal and ordinal measurement is dealt with in the next section on regulatory compliance/licensing and program quality).

9) QRIS versus Licensing: examples of a differential/relative regulatory compliance monitoring system would be QRIS – Quality Rating and Improvement Systems. Absolute/full regulatory compliance systems would be state licensing systems. Many programs talk about the punitive aspects of the present human services licensing and monitoring system and its lack of focus on the program quality aspects in local programs. One should not be surprised by this because in any regulatory compliance system the focus is on "doing no harm" rather than "doing things well". It has been and continues to be the focus of licensing and regulations in the USA. The reason QRIS - Quality Rating and Improvement Systems developed in early care and education was to focus more on "doing things well" rather than "doing no harm". This is not the case in many Canadian Provinces and European countries in which they have incorporated program quality along with specific regulatory requirements.

10) Non-Linear versus Linear: the assumption in both differential/relative and absolute/full regulatory compliance monitoring systems is that the data are linear in nature which means that as compliance with rules increases positive outcomes for clients increases as well. The problem is the empirical data does not support this conclusion. It appears from the data that the relationship is more non-linear where there is a plateau effect with regulatory compliance in which client outcomes increase until substantial compliance is reached but doesn't continue to increase beyond this level. There appears to be a "sweet spot" or balancing of key rules that predict client outcomes more effectively than 100% or full compliance with all rules – this is the essence of the Theory of Regulatory Compliance (Fiene, 2019) – substantial compliance with all standards or full compliance with a select group of standards that predict overall substantial compliance and/or positive client outcomes.

As the regulatory science and administrative fields in general continue to think about the appropriate monitoring systems to be designed and implemented, the above structure should

help in thinking through what these measurement systems' key elements should be. Both paradigms are important, contexts, but a proper balance between the two is probably the best approach in designing regulatory compliance monitoring systems.

Regulatory Compliance/Licensing and Quality

This part of the policy commentary will delineate the differences between regulatory compliance and quality. It will provide the essential principles and elements that clearly demonstrate the differences and their potential impact on program monitoring. Obviously, there is some overlap between this section and the above section dealing with regulatory compliance monitoring paradigms. When we think about regulatory compliance measurement, we are discussing licensing systems. When we think about quality, we are discussing Quality Rating and Improvement Systems (QRIS), accreditation, professional development, or one of the myriad quality assessment tools, such as the Classroom Assessment Scoring System (CLASS) or Environment Rating Scales (ERS's). All these systems have been designed to help improve the health and safety of programs (licensing) to building more environmental quality (ERS), positive interactions amongst teachers and children (CLASS), enhancing quality standards (QRIS, accreditation), or enhancing teacher skills (professional development).

There are eight basic principles or elements to be presented (they are presented in a binary fashion demonstrating differences): 1) "Do no harm" versus "Do good". 2) Closed system versus Open system. 3) Rules versus Indicators. 4) Nominal versus Ordinal measurement. 5) Full versus Partial compliance. 6) Ceiling effect versus No Ceiling effect. 7) Gatekeeper versus Enabler. 8) Risk versus Performance.

1) Let's start with the first principal element building off what was discussed in the above section, "Do No Harm" versus "Do Good". In licensing, the philosophy is to do no harm, its emphasis is on prevention, to reduce risk to children in a particular setting. There is a good deal of emphasis on health and safety and not so much on developmentally appropriate programming. In the quality systems, such as QRIS, accreditation, professional development, Environmental Rating Scales, CLASS, the philosophy is to do good, its emphasis is looking at all the positive aspects of a setting. There is a good deal of emphasis on improving the programming that the children are exposed to or increasing the skill set of teachers or improving the overall environment or interaction that children are exposed to.

2) Closed system versus Open system. Licensing is basically a closed system. It has an upper limit with full compliance (100%) with all rules. The goal is to have all programs fully comply with all rules. However, the value of this assumption has been challenged over the years with the introduction of the Regulatory Compliance Theory of Diminishing Returns (Fiene, 2019). With quality systems, they tend to be more open and far reaching where attaining a perfect score is very difficult to come by. The majority of programs are more normally distributed where with licensing rules

the majority of programs are skewed positively in either substantial or full compliance. It is far more difficult to distinguish between the best programs and the mediocre programs within licensing but more successful in quality systems.

3) Rules versus Indicators/Best Practices. Licensing systems are based around specific standards/rules/regulations that either are in compliance or out of compliance. It is either a program is in compliance or out of compliance with the specific rule. With quality systems, there is more emphasis on indicators or best practices that are measured a bit more broadly and deal more with process than structure which is the case with licensing. It is the difference between hard and soft data as many legal counsels term it. There is greater flexibility in quality systems. With this said, if we can look at other service types, such as adult-residential services, there has been some limited success with blending structural and process elements but it still remains a measurement issue on the process side.

4) Nominal versus Ordinal measurement[3]. Licensing systems are nominally based measurement systems. Either you are in compliance or out of compliance. Nothing in-between. It is either a yes or no response for each rule. No maybe or partial compliance. With quality systems, they are generally measured on an ordinal level or a Likert scale. They may run from 1 to 3, or 1 to 5, or 1 to 7. There are more chances for variability in the data than in licensing which has 1 or 0 response. This increases the robustness of the data distribution with ordinal measurement.

5) Full or None versus Gradients or Gray Area. Building off of the fourth element, licensing scoring is either full or not. As suggested in the above elements, there is no in-between category, no gradient or gray area. This is definitely not the case with quality systems in which there are gradients and substantial gray areas. Each best practice can be measured on a Likert scale with subtle gradients in improving the overall practice.

6) Ceiling effect versus No Ceiling. With licensing there is definitely a ceiling effect because of the emphasis on full 100% compliance with all rules. That is the goal of a licensing program, to have full compliance. With quality systems, it is more open ended in which a ceiling effect is not present. Programs have many ways to attain excellence.

7) Gatekeeper versus Enabler: Licensing has always been called a gatekeeper system. It is the entry way to providing care, to providing services. It is a mandatory system in which all programs need to be licensed to operate. In Quality systems, these are voluntary systems. A program chooses to participate, there is no mandate to participate. It is more enabling for programs building upon successes. There are enhancements in many cases.

8) Risk versus Performance: Licensing systems are based upon mitigating or reducing risks to children when in out of home care. Quality systems are based upon performance and excellence where this is rewarded in their particular scoring by the addition of a new Star level or a Digital Badge or an Accreditation Certificate.

There has been a great deal of discussion in the early care and education field about the relationship between licensing, accreditation, QRIS, professional development, and technical

assistance. It is important as we continue this discussion to pay attention to the key elements and principles in how licensing and these quality systems are the same and different in their emphases and goals, and about the implications of particular program monitoring paradigms and measurement strategies. For other regulatory systems outside the human services field, the same type of model can be applied positioning compliance and quality as a continuum one building off of the other because I feel that with the introduction of more quality into a regulatory context will help to ameliorate the ceiling and plateau effect of diminishing returns on performance and outcomes.

Reference:

Fiene, R. (2019). A Treatise on Regulatory Compliance. *Journal of Regulatory Science*, Volume 7, 2019

Notes:

1. This manuscript should be read along with *A Treatise on Regulatory Compliance* which is referenced above because the two articles build off one another. In the *treatise* description, the specific idiosyncrasies of regulatory compliance data and other key implications of the theory are pointed out that enhance the presentation in this article, such as the extreme nature of skewness that is present in regulatory compliance data, nominal data measurement, the differences between full and substantial regulatory compliance, designing the most cost effective and efficient differential monitoring system, and the need to dichotomize data because of the skewed nature of the data distribution.

2. The ultimate goal is the most cost effective and efficient differential monitoring system for negotiating the path of least resistance in complying with a given set of rules which will provide the proper balance of rules. This should be the goal of any regulatory science paradigm. By using the previous *Treatise* article along with this article should provide a blueprint for the regulatory science field in designing a program monitoring system to measure regulatory compliance where an emphasis on differential monitoring should occur in licensing systems and full-scale monitoring should occur in program quality systems. Another approach is to have both regulatory compliance and program quality built as a continuum in the program monitoring system similar to what Head Start is attempting.

3. There are instances in which this dichotomy is not as clear or straightforward where licensing systems do allow partial compliance as a facility has opportunities to correct non-compliances on their way to achieving full compliance with specific rules. The problem is that this is not necessarily a standardized process and it is difficult to determine if it is used often in licensing agencies' monitoring efforts.

International Journal of Child Care and Education Policy
2013, Vol. 7, No. 1, 1-15
Copyright 2013 by Korea Institute of Child Care and Education

A Comparison of International Child Care and US Child Care Using the Child Care Aware – NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks

Richard Fiene, Ph.D.

Affiliated Faculty
Prevention Research Center
The Pennsylvania State University

This is a first of its kind study comparing the USA to other world countries utilizing the Child Care Aware – NACCRRA Child Care Benchmarks related to health and safety rules and regulations. A team of researchers analyzed the child care/early care & education rules and regulations from the USA and a selected group of countries to do a comparative analysis using the Child Care Aware – NACCRRA benchmarking scoring protocol. The results from the analyses were somewhat unexpected in that the scores between the USA and the other countries were not as statistically significant in the overall scores. However, when more specific benchmarks were compared statistically significant differences did appear in the health & safety and professional development areas.

Key words: Child Care Quality, Comparisons of USA and International Child Care, Child Care Regulations.

Introduction

The purpose of this paper is to compare several countries (N =20) and the United States on the Child Care Aware – formerly NACCRRA (National Association of Child Care Resource and Referral Agencies) Child Care Benchmarks that have used extensively in the USA to compare state regulatory and monitoring policy and implementation. The use of these benchmarks has been very useful in comparing states in the USA on an agreed upon series of child care benchmarks that have a great deal of support in the research literature (AAP/APHA, 2012, 2013; NACCRRA 2007, 2009, 2011). Previous research (OCED, 2006) has focused on early care and education policies in other countries which was a very important

41 Grandview Avenue
Middletown, Pennsylvania USA 17057
717-944-5868 Phone and Fax
Fiene@psu.edu
March 2013 (revised and resubmitted July 2013)

first step in making comparisons across countries. This paper will expand upon this comparison in order to begin applying the NACCRRA benchmarks to other countries and establish a baseline between the USA and other countries related to regulatory review and analysis. This study is important because it provides a common rubric for making comparisons between the USA and other countries that is reliable and valid (NACCRRA 2007, 2009, 2011) related to regulatory analysis. As far as the author can determine from his extensive review of the literature, similar studies of this type have not been attempted utilizing a standardized rubric created by a major national child care organization. There have been other studies completed in which comparisons were made of other countries, the OCED (2006) Starting Strong II study and report is an excellent example of this type of

DIFFERENTIAL MONITORING LOGIC MODEL & ALGORITHM (DMLMA©) (Fiene, 2012): A 4th Generation ECPQIM – Early Childhood Program Quality Indicator Model

CI x PQ => RA + KI => DM

Definitions of Key Elements:

CI = Comprehensive Licensing Tool (Health and Safety)(***Caring for Our Children***)
PQ = ***ECERS-R, FDCRS-R, CLASS, CDPES*** (Caregiver/Child Interactions/Classroom Environment)
RA = Risk Assessment, (High Risk Rules)(***Stepping Stones***)
KI = Key Indicators (Predictor Rules)*(13 Key Indicators of Quality Child Care)*(NACCRRA Benchmarks)
DM = Differential Monitoring (How often to visit and what to review)

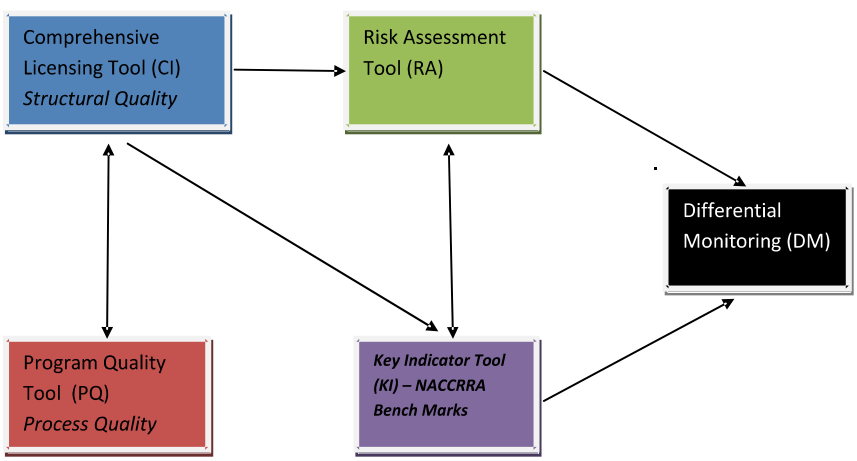

Comprehensive Licensing Tool (CI) Structural Quality
Risk Assessment Tool (RA)
Differential Monitoring (DM)
Program Quality Tool (PQ) Process Quality
Key Indicator Tool (KI) – NACCRRA Bench Marks

Figure 1.

analysis and is recommended reading for anyone interested in reviewing public policy analyses.

The child care benchmarks[1] utilized in this study are based upon the following key indicators: prevention of child abuse, immunizations, staff child ratio, group size, staff qualifications and training, supervision/discipline, fire drills, medication administration, emergency plan/contact, outdoor playground, inaccessibility of toxic substances, and proper hand washing/ diapering (NACCRRA 2007, 2009, 2011). These benchmarks are more based upon the structural aspects of quality rather than on the process aspects of quality. This is an important distinction between the USA approach and the other countries approaches that becomes important in the explanation of results later in this paper.

This paper also supports and expands the development of an Early Childhood Program Quality Indicator Model (ECPQIM)(Fiene & Nixon, 1985) which is in a 4th generation (Fiene, 2013) as a differential monitoring logic model & algorithm helping to guide the program monitoring of child care/early care & education programs (see Figure 1).

Method

Data Collection Process

Data collection was done on a 100 point scale which is delineated in Appendix 1 as developed by the Child Care Aware - NACCRRA Research Team. The same scoring protocol that was utilized in developing the 2007, 2009, and 2011 Reports and comparisons of states by Child Care Aware - NACCRRA was employed in this study in comparing the average scores of the states and the 20 countries. The 100 point scale consisted of 10 child care benchmarks each worth 10 points: ACR = Staff child ratios NAEYC Accreditation Standards met (R1); GS = Group size NAEYC Accreditation Standards met (R2); Director = Directors have bachelor's degree (R3); Teacher = Lead teacher has CDA or Associate degree (R4); Pre = Initial orientation training (R5); Inservice = 24 hours of ongoing training (R6); Clearance = Background check (R7); Devel = Six developmental domains (R8); Health = Health and safety recommendations (R9); and Parents = Parent Involvement (R10).

Data Scoring

The scoring protocol employed a total raw score approach of 100 points that was used to compare the countries on the 10 child care benchmarks in the aggregate. The scoring protocol also employed a standardized scoring approach (0 to 2 points) on each of the 10 child care benchmarks utilizing the following scale: 0.0 = Does not meet the Child Care Aware – NACCRRA Benchmarks; 0.5 = Marginally meets the Child Care Aware – NACCRRA Benchmarks; 1.0 = Partially meets the Child Care Aware – NACCRRA Benchmarks; 1.5 = Substantially meets the Child Care Aware – NACCRRA

Benchmarks; 2.0 = Fully meets the Child Care Aware – NACCRRA Benchmarks.

Data Collectors

A team of undergraduate and graduate research assistants[2] at the Pennsylvania State University were the data collectors in which each of them reviewed the child care/early childhood rules/regulations/standards from a specific country and scored the rules/regulations/standards on the Child Care Aware – NACCRRA 100 point raw score protocol and the standardized (0 – 2) scoring approach.

Data Sources

The child care regulations selected were for preschool age children only in child care center setting in the 20 countries. Geographically the governmental jurisdiction closest to the national capital was used if applicable national regulations could not be found. More than the final 20 countries selected were reviewed but several countries needed to be dropped because they did not meet the above criteria or the regulations could not be found in English. This was more a convenience sample rather than a stratified scientific sample, a limitation of this study.

Results

The results from this study and analysis were totally unexpected. The results indicated no statistically significant differences between the USA and the other countries selected (Australia, Belgium, Norway, Finland, Sweden, Ireland, United Kingdom, Italy, France, New Zealand, Mexico, Greece, Canada, Austria, Portugal, Philippines, Turkey, Pakistan, Nigeria, Denmark, and Spain – these countries were selected because of their availability of child care/early care & education rules and regulations as described previously above in Data Sources) when comparing the total scores on the 100 point scale; the USA average for all 50 states scored 58 while the 20 countries average score was 56. However, a very different scenario occurs when looking at the ten individual child care benchmarks using the standardized 0 – 2 scoring protocol. The 20 countries selected in this study scored statistically higher on the following child care benchmarks: Director ($t = 7.100$; $p < .0001$) and Teacher ($t = 7.632$; $p < .0001$) qualifications. The USA scored statistically higher on the following child care benchmarks: Health/Safety ($t = 6.157$; $p < .0001$), Staff Clearances ($t = 3.705$; $p < .01$), and Pre-Service ($t = 4.989$; $p < .001$) /In-Service training ($t = 2.534$; $p < .02$) (See Table 1 & Figure 2).

The results showed that both the USA and all other countries mean scores were 58 and 56 respectively on the 100 point scale – this is a raw scale score and not the standardized score (0 – 2 – see Table 1 and Figure 2) which was used in the comparisons for each benchmark. This is not a particularly good score if you think in terms of exams, but for states and countries with

Table 1

Mean Comparisons between USA and Twenty Countries on Child Care Aware – NACCRRA Benchmarks

Benchmark	Countries	USA	Significance
ACR (R1)	1.122	0.8462	not significant
GS (R2)	0.4063	0.5865	not significant
Director (R3)	**1.5625**	0.5	t = 7.100; p < .0001
Teacher (R4)	**1.6563**	0.4038	t = 7.632; p < .0001
Preservice (R5)	0.9375	**1.6731**	t = 4.989; p < .001
Inservice (R6)	0.6563	**1.0481**	t = 2.534; p < .02
Clearances (R7)	0.6094	**1.2404**	t = 3.705; p < .01
Development (R8)	1.6406	1.4519	not significant
Health(R9)	0.9844	**1.7404**	t = 6.157; p < .0001
Parent(R10)	1.5000	1.5385	not significant

Legend:
Child Care Aware - NACCRRA Benchmarks:
Parent = Parent Involvement (R10)
Health = Health and safety recommendations (R9)
Development = Six developmental domains (R8)
Clearances = Background check (R7)
Inservice = 24 hours of ongoing training (R6)
Preservice = Initial orientation training (R5)
Teacher = Lead teacher has CDA or Associate degree (R4)
Director = Directors have bachelor's degree (R3)
GS = Group size NAEYC Accreditation Standards met (R2)
ACR = Staff child ratios NAEYC Accreditation Standards met (R1)

Scoring:
0.0 = Does not meet Child Care Aware – NACCRRA Benchmarks.
0.5 = Marginally meets Child Care Aware – NACCRRA Benchmarks.
1.0 = Partially meets Child Care Aware – NACCRRA Benchmarks.
1.5 = Substantially meets Child Care Aware – NACCRRA Benchmarks.
2.0 = Fully meets Child Care Aware – NACCRRA Benchmarks.

vastly complex bureaucracies maybe this isn't as bad as it looks. Could it be that the USA is better than we think or is it that the USA and all other countries are providing just mediocre child care?!

The reason for using aggregate data in this study was to be consistent in how data have been collected in the USA utilizing the Child Care Aware – NACCRRA Scoring Protocol. This did delimit the potential analyses for this study and the recommendation would be made in future studies to unbundle

the results so that more detailed comparisons could be made. As mentioned in the introduction, the purpose of this study was to provide an initial baseline comparison between the USA and other countries on the Child Care Aware – NACCRRA Scoring Protocol.

Discussion

The purpose of this study was to extend the Child Care Aware - NACCRRA Child Care Benchmarks Scoring Protocol to an international sample comparison. As has been done by the National Science Foundation with math and science testing, these same types of comparisons have been made with the USA not fairing all that well on the math and science comparisons.

It appears that when it comes to child care benchmarks the USA actually appears to be in better shape than many advocates and experts would have thought when compared to other countries or is it that the other countries are providing the same form of mediocre care as it relates to these child care benchmarks. Remember that these benchmarks are heavily weighted towards the structural side of quality

Legend:
Child Care Aware - NACCRRA Benchmarks:
Parents = Parent Involvement (R10)
Health = Health and safety recommendations (R9)
Devel = Six developmental domains (R8)
Clearance = Background check (R7)
Inservice = 24 hours of ongoing training (R6)
Pre = Initial orientation training (R5)
Teacher = Lead teacher has CDA or Associate degree (R4)
Director = Directors have bachelor's degree (R3)
GS = Group size NAEYC Accreditation Standards met (R2)
ACR = Staff child ratios NAEYC Accreditation Standards met (R1)

Scoring:
0.0 = Does not meet Child Care Aware – NACCRRA Benchmarks.
0.5 = Marginally meets Child Care Aware – NACCRRA Benchmarks.
1.0 = Partially meets Child Care Aware – NACCRRA Benchmarks.
1.5 = Substantially meets Child Care Aware – NACCRRA Benchmarks.
2.0 = Fully meets Child Care Aware – NACCRRA Benchmarks.

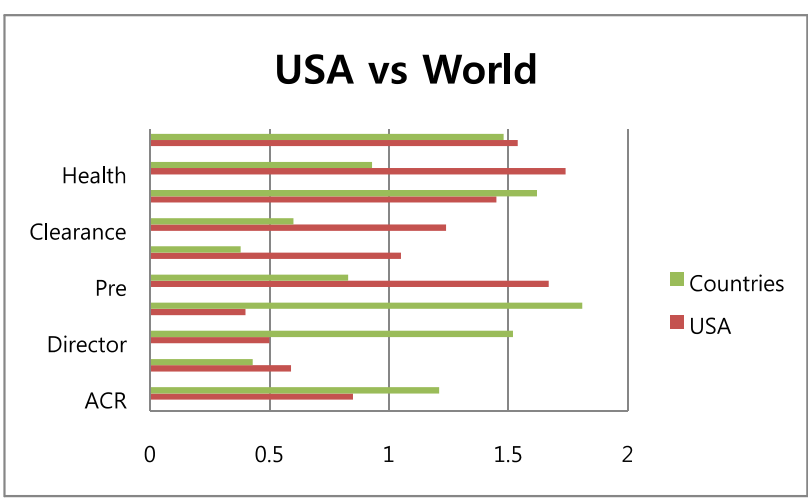


Figure 1. *Mean Comparisons between USA and Twenty Countries on Child Care Aware – NACCRRA Benchmarks*

rather than the process side of quality.

However, when the individual benchmarks are analyzed then certain patterns occur which seem very consistent with the previous research literature. The 20 countries scored higher on the staffing benchmarks while the USA scored higher on the training and health/safety benchmarks. Clearly this is an indication reflecting public policy in the other countries as versus the USA. Many other countries place more emphasis on the process aspects of quality which involve staff and staff interactions with children. The USA has focused more on the structural aspects of quality which involve health & safety especially in the state licensing of child care. These structural aspects of quality are more easily quantifiable in state rules and regulations which is the locus of control for the licensing of child care. Since the USA does not have national standards that are required (the USA does have national health and safety standards that are recommended practice, such as Caring for Our Children (2012)) as is the case in so many of the countries in this study, this may provide a possible explanation for the results of this study. It will be interesting to see how Quality Rating and Improvement Systems (QRIS) which usually have some process standards impact this overall balance of structural and process aspects of quality. This is an area that needs additional research and more in-depth analysis.

So what does this tell us. I think it is a warning call as has been put forth by Child Care Aware - NACCRRA that we still have a lot of additional work to do in improving child care, not only in the USA, but worldwide. Just as the Child Care Aware -NACCRRA Report Cards (2007, 2009, 2011) have played a role in making positive change in the child care benchmarks over time; we need to expand this reporting and change to a world wide focus. There is clearly the need to expand from the present analysis of 20 countries and the USA to other countries throughout the world and to track changes over time as Child Care Aware/NACCRRA has done.

Another area of concern within the USA and I am sure in other countries as economies have begun their slow recovery from the economic downturn of 2008 – 2010 is to do more with less. One such approach being explored in the USA is called differential monitoring which helps to re-allocate limited resources in a more cost effective and efficient manner via a risk assessment and key indicator approach. I hope that this comparison utilizing the Child Care Aware – NACCRRA Benchmarking Scoring Protocol and introducing the Early Childhood Program Quality Indicator Model/Differential Monitoring Logic Model and Algorithm (Fiene, 2013) within an international context as first steps in making that happen.

References

AAP/APHA (2012). *Caring for our children*, Washington, D.C.: American Public Health Association.

AAP/APHA (2013). *Stepping Stones*, Washington, D.C.: American Public Health Association.

Child Care Award - NACCRRA, (2007). *We Can Do Better: NACCRRA's Ranking of State Child Care Center Standards and Oversight*, Washington, D.C.: National Association of Child Care Resource and Referral Agencies.

Child Care Aware - NACCRRA, (2009). *We Can Do Better: 2009 Update: NACCRRA's Ranking of State Child Care Center Regulations and Oversight*, Washington, D.C.: National Association of Child Care Resource and Referral Agencies.

Child Care Aware - NACCRRA, (2011). *Leaving Children to Chance: NACCRRA's Ranking of State Standards and Oversight in Child Care Centers*, Washington, D.C.: National Association of Child Care Resource and Referral Agencies

Fiene, R. (2013). *Differential monitoring logic model and algorithm*. Middletown, Pennsylvania: Research Institute for Key Indicators.

Fiene, R. & Nixon, M. (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly, 14*(3), 198-214.

OCED (2006). *Starting strong II*. Paris, France: Organization for Economic Co-Operation and Development Publishing.

Notes

[1] In the licensing literature these child care benchmarks are usually referred to as key indicators (Fiene, 2013). Please see Figure 1 which delineates where within a program monitoring system these benchmarks would appear and could be utilized.

[2] The following individuals played key data collection roles as research assistants in the compilation of this study: Melissa Cave, Ashley Le, Breanna Green, Corrie Podschlne, Sherrie Laporta, Ashley Edwards, Laura Hartranft, Gissell Reyes, Janet Lazur, Kayma Freeman, Jessica White, Karen Mapp, and Lindsay Bitler.

Appendix 1

Benchmark criteria for *We Can Do Better: NACCRRA Ranking of State Child Care Center Regulations:2011 Update* were developed by Child Care Aware - NACCRRA and have been used for the 2007, 2009 and 2011 We Can Do Better reports. The rationale for each standard, including research evidence of its importance in quality care, is noted in each section of the report and in previous reports. Each of the 10 regulation benchmarks were scored with a value ranging from one to 10 points, depending on how closely the state met the benchmark, for a maximum total of 100 points. In cases where states permit several different options for complying (e.g., complying with director or teacher qualifications), the minimum allowed was used. This information was used to generate state sheets with scores for each standard.

Scoring Methods for NACCRRA Ranking of State Child Care Center Regulations (R)

Question	Scoring method
Regulation 1. Staff:child ratio requirements comply with NAEYC accreditation standards.	

6 mo	9 mo	18 mo	27 mo	3 yr	4 yr	5 yr
1:4	*1:4*	*1:4*	*1:4*	*1:9*	*1:10*	*1:10*

Number of ratios in compliance with NAEYC standards	Score
7 ratios	10
6 ratios	9
5 ratios	8
4 ratios	7
3 ratios	5
2 ratios	3
1 ratios	1

R2. Group size requirements are in compliance with NAEYC accreditation standards.

6 mo	9 mo	18 mo	27 mo	3 yr	4 yr	5 yr
8	*8*	*8*	*8*	*18*	*20*	*20*

Number of group sizes in compliance with NAEYC standards	Score
7 ratios	10
6 ratios	9
5 ratios	8
4 ratios	7
3 ratios	5
2 ratios	3
1 ratios	1

R3. Center directors are required to have a bachelor's degree of higher in early childhood education or a related field.

Director education requirement	Score
Bachelor's degree in any field	10
College directors certification	7
Any associate degree	5
CDA	5
Clock hours/less than associate degree	2
High school or less	0

R4. Lead teachers are required to have a Child Development Associate (CDA) credential or an associate degree in early childhood education or related field.

Lead teacher education requirement	Score
CDA/associate degree or better	10
State Credential	5
Clock Hours in ECE	2
High School/GED	2
Less than High School	0

R5. Lead teachers are required to have initial training, including:

- Orientation.
- Fire safety.
- Other health and safety issues.
- At least one staff member certified in first aid must be present when children are in care.
- At least one staff member who is certified in CPR must be present when children are in care.

Number of areas training is required	Score
Five areas	10
Four areas	8
Three areas	6
Two areas	4
One area	2
None	0

R6. Lead teachers are required to have 24 hours or more of annual training.

Ongoing training ≥	Score
24 Hours	10
18 hours	7
12 hours	5
6 hours	2
None	0

R7. A comprehensive background check is required for child care providers.

- Use of fingerprints to check state records.
- Check FBI records.
- Check state child abuse registry
- Check sex offender registry.
- Criminal history check.

Number of Background checks completed	Score
Five checks	10
Four checks	8
Three checks	6
Two checks	4
One check	2
None	0

R8. Child care centers are required to offer program activities that address all six child development domains

- Language/literacy.
- Cognitive.
- Social.
- Emotional.
- Physical.
- Cultural.

Developmental domains addressed	Score
6 domains	10
5 domains	9
4 domains	7
3 domains	5
2 domains	3
1 domain	1
None	0

R9. Child care centers are required to follow 10 recommended health and safety practices.

- Immunizations.
- Guidance/discipline.
- Diapering and handwashing.
- Fire drills.
- Medication administration.
- SIDS prevention.
- Emergency preparedness.
- Playground surfaces.
- Hazardous materials.
- Incidence reporting.

Standards addressed	Score	Standards addressed	Score
10	10	5	5
9	9	4	4
8	8	3	3
7	7	2	2
6	6	1	1
Allowing corporal punishment is an automatic zero			

R10. Child care centers are required to:

- Encourage parent involvement.
- Require daily or ongoing communication with parents.
- Allow parental access any time their children are in care.

Number of items required	Score
Three items	10
Two items	7
One item	3
None	0

Appendix 2

These were the countries included in these analyses: Australia, Belgium, Norway, Finland, Sweden, Ireland, United Kingdom, Italy, France, New Zealand, Mexico, Greece, Canada, Austria, Portugal, Philippines, Turkey, Pakistan, Nigeria, Denmark, Spain, and the USA which included all 50 states.

The Instrument Based Program Monitoring Information System and the Indicator Checklist for Child Care

Richard Fiene
Office of Children Youth and Families
Commonwealth of Pennsylvania
Mark Nixon
Children's Services Monitoring Transfer Consortium, Washington, D.C.

ABSTRACT: The Instrument Based Program Monitoring Information System (IPM) and the Indicator Checklist (IC) are two tools for the state management of child day care services. A methodology for monitoring interviews and site visits to child day care programs is described. An integral feature of IPM is a system of assigning weights to the questions or items so that scores reflect the relative importance of state regulations. An Indicator Checklist is a questionnaire or checklist that contains selected, predictive items from a longer, comprehensive instrument that a state uses to monitor child day care providers' conformance to state day care regulations. An Indicator Checklist contains items that have been determined to be most effective in discriminating between providers that typically receive high overall scores on the comprehensive instrument and providers that typically receive low overall scores.

For nearly half a century, state governments have accepted responsibility for ensuring that those who care for children in their home and in day care centers meet minimum requirements for health and safety. During the past decade as the amount of state and federal funds for day care have grown, states have taken an active role in monitoring (1) the ways in which day care providers administer their programs, and (2) the quality of the services provided to children for whose care the state is paying.

Nationally, day care is big business. It is estimated that currently there are more than 118,000 licensed providers who serve an estimated 1.2 million children every day. The stakes in assuring that these children are well served are high, both in terms of public health and safety and from the viewpoint of enhancing the growth and development of America's most precious resource, its children. It is estimated that $6.3 billion dollars are spent annually on day care services.[1]

Reprints should be requested from Richard Fiene, Directory of Research and Information Systems, Office of Children, Youth, and Families, 1514 North Second Street, Harrisburg, PA 17102.

[1] Day care services include group day care centers serving 12 or more children, group day care homes serving 6–11 children, and family day care homes serving 5 or fewer children. Head Start & nursery school programs that operate for part day are included in day care services definition.

Child Care Quarterly, 14(3), Fall 1985
©1985 by Human Sciences Press

However, in monitoring these services, states spend less than one percent of their day care funds each year to ensure that providers comply with regulations or meet quality guidelines.

This article describes an approach in monitoring child day care services called: Instrument Based Program Monitoring (IPM). An IPM differs substantially from the more common approach to monitoring: narrative site visit reports used by most states. The narrative report approach usually includes a site visit to each provider and the preparation of a summary of observations and interpretive and evaluative comments about the monitor's findings. These reports are time consuming to prepare, and often difficult to summarize succinctly for policy makers and administrators. This article describes an alternative to the narrative site report.

Forces Changing the Regulatory Environment

The job of state agencies in program monitoring is currently changing in response to powerful forces in American society, especially at the level of state government.

First, there is the continuing need to assure parents that their children will not be subjected to unsafe day care environments and that day care providers who receive state funds are meeting the terms of their contracts with the state by providing quality services. Quality services are defined as day care services that promote sound child development principles and do not only ensure that children are in healthy and safe child care environments. Public accountability requires that the state entertain a dual purpose, one is to monitor compliance with state regulations; but secondly and equally important, there is a strong need for the state to ensure that quality child development services are supported and provided.

Gwen Morgan's (1980) work is particularly helpful in providing direction regarding the relationship between licensing and funding criteria. A Model presented by Morgan (1980) clearly delineates a regulatory continuum where day care licensing is considered as the floor to quality with accreditation as the standard of quality for which model day care programs strive. Recent efforts by the National Association for the Education of Young Children (Center Accreditation Project (1983)) and the Children's Services Monitoring Consortium (Child Development Program Evaluation Scale (1984)) have helped to support this move towards accreditation and the measurement of quality in early childhood programs. These efforts take on additional meaning given the direction from the federal government to pass as much of the responsibility for monitoring early childhood programs to the states.

Second, the fiscal cutbacks that are now occurring in many states will almost certainly increase the pressure on state agencies to operate as efficiently as possible. Cutbacks in staff across agencies are likely, even as workloads increase. These factors will force states to streamline their regulatory enforcement and monitoring efforts in all areas, including day care and children's services. A promising approach attempted in some states is moving from a licensing to a registration system. In a registration system, the locus of control for the regulatory process is shifted from the state to the provider level—the provider is responsible for assuring that s/he meets all registration requirements.

Third, the role of the state in regulating private sector organizations is changing. There are now active pressures to reduce the general level of state regulation with a view toward encouraging private market forces in the production and allocation of goods and services. Further, there is a commitment in a growing number of states to reduce the extent of the Federal Government's involvement, including federal funding and accompanying regulatory requirements, in several areas, notably human services (The moratorium placed on the Federal Interagency Day Care Requirements is a specific example which was supported by a number of states).

Fourth, many states are actively seeking ways to reduce the burden on the private sector of the compliance monitoring activities that are perfomed by the state. For those regulations that continue in force, many states will be examining approaches that simplify monitoring procedures and make them less onerous for providers. This is particularly true for day care services, which are often provided by individuals or organizations that may have little experience coping with regulations.

IPM as a Response to These Forces

One approach that states have used to cope with these forces is the development of Instrument-Based Program Monitoring Systems—(IPMs).

As the name implies, an IPM system incorporates three distinguishing characteristics: *First,* it is instrument-based. The system uses checklists or questionnaires that contain highly specific questions. These questions usually correspond directly to the state's regulations or other requirements (e.g., fiscal requirements). *Second,* it supports program monitoring. In its broadest sense, program monitoring is the management process of conducting periodic reviews

or inspections to ensure that certain activities, such as the provision of day care service, meet acceptable criteria, and the process of effecting corrective action where required. Program monitoring may include one or some conbination of:

1. Licensing reviews (Table 1 gives a listing of items taken from Pennsylvania's IPM at the licensing and minimal standards level);
2. Contract compliance reviews; and
3. Evaluations of program quality that go beyond minimum requirements to health and safety. A specific example that may be helpful is taken from the *California Child Development Program Quality Review* (1982) Instrument. What follows is a sampling of the Table of Contents:

PROGRAM QUALITY SUB SCALE

A. GOALS AND OBJECTIVES OF CHILD DEVELOPMENT PROGRAM ARE EVALUATED AT LEAST ANNUALLY BY THE STAFF AND PARENTS AND ARE MODIFIED AS NEEDED
B. TEACHING STAFF HIGHLIGHTS EACH CHILD BY SHARING INDIVIDUAL ETHNIC AND CULTURAL BACKGROUNDS—EMPHASIS IS PLACED ON CARE-GIVER OBSERVATIONS.
C. THE GOALS, OBJECTIVES, AND PROCEDURE FOR IDENTIFICATION OF CHILDREN'S NEEDS ARE EVALUATED AT LEAST ANNUALLY BY STAFF AND PARENTS (Fiene, 1984).

Third, IPM is a comprehensive system. It is part of a group of related steps such as on-site reviews, corrective action, follow-up reviews, and summarizing and reporting results that are used recurrently to accomplish the task of compliance monitoring. Program, fiscal, and statistical components can be linked quantitatively to constitute a comprehensive IPM system for day care. A new software decision support system (Watson, Fiene, & Woods, 1984) based on IPM is being developed for micro-computer technology and is being pilot tested in Michigan Department of Social Services, and Texas Department of Human Resources. When the IPM system is used in this linked fashion, it provides the basis for monitoring child day care Vendor & Voucher Delivery systems.

The advantages of an IPM system that are responsive to the changes mentioned earlier include: consistency, coverage of all regulatory areas, clear expectations simplified monitoring procedures,

TABLE 1

Pennsylvania Child Development Program Evaluation

Specific Items Within Identified General Areas

General Requirements	
1. Relevant approvals	4. Child abuse reporting procedures
2. Insurance coverage	5. Provision for special services
3. Parent participation	
Staffing Standards	
1. Qualifications of staff	staff requirements
2. Responsibilities	4. Staff health requirements
3. Adult/child ratio and minimum	
Employee Records	
1. Evidence of qualifications and references for staff	
Building & Site	
1. Appropriate indoor and outdoor square footage per child	materials
2. Characteristics of play areas	5. Cleanliness
3. Sanitary facilities	6. Screening of windows and doors
4. Storage of medicine and	7. Heating apparatus
	8. Educational materials available
Equipment	
1. Condition and placement of equipment	2. Swimming regulations
	3. Napping rules
Program for Children	
1. Evidence of written program plan with developmental activities	special needs children
2. Discipline	4. Sanitary habits developed
3. Identification and referral of	5. Infant/toddler stimulation
	6. School-age requirements
Food & Nutrition	
1. Menu requirements	3. Utensils
2. Infant formula rules	4. Special diet considerations
Transportation	
1. Vehicles all licensed and inspected	4. Restraint of children
2. Insurance coverage	5. First-aid kit materials
3. Adult/child ratio	
Child Health	
1. Requirements of health records	4. Medications
2. Emergency contact information	5. Procedure for ill children
3. Medical emergency procedures	6. First-aid requirements
Staff Health	
1. Procedures for staff illness	2. Physical requirements for infant caregivers
Procedures & Applications	
1. Pre-admission policy	3. Requirements of day care agreement
2. Requirements for child's application	
Child Records	
1. Frequency of updating records	4. Parental rights to records
2. Confidentiality	5. Procedure for release of information
3. Information to be included in child's records	6. Use of records after termination of service

and potential for cost efficiencies. With an IPM system, the same questionnaire or checklist is used with all providers, and there is less opportunity for individual bias in reporting results. Similarly, basing the questions or checklist items explicitly on the regulations or other requirements makes it possible to ensure that all areas are covered adequately. Having a clear set of questions that are known to both monitoring staff and providers reduces the possibility of misunderstandings and misinterpretations concerning the results of the review. Finally, standardized procedures for administering the questionnaire and processing the results can simplify the state's monitoring task and reduce the time, cost, and burden of monitoring both to the provider and to the state.

Four agencies (Pennsylvania's Office of Children Youth and Families, West Virginia's Office of Social Services, California's Office of Child Development, and New York City's Agency for Child Development) that are part of a consortium for improving the monitoring of children's services (Children's Services Monitoring Transfer Consortium) have experienced significant improvements in provider satisfaction with monitoring efforts and have, in some cases, achieved more efficient allocations of resources for day care and day monitoring. Pennsylvania has experienced substantial cost savings by linking the results of their IPM system to the state's fiscal and statistical information systems (See Figure 1). The state was able to set a ceiling on

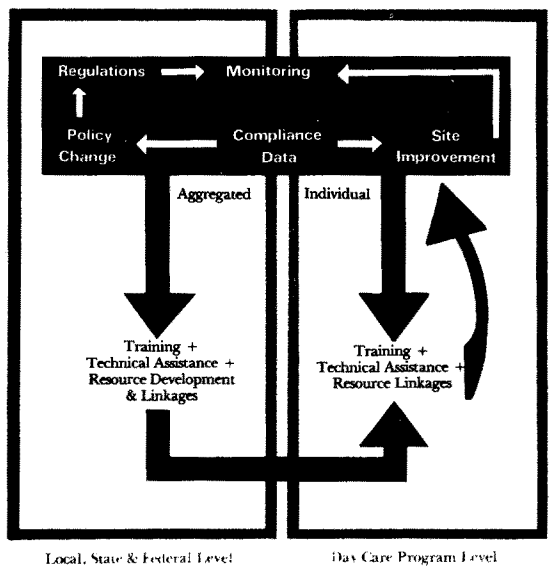


FIGURE 1

day care funding that did not jeopardize program quality, and used the funds that were formerly given to high-cost providers to improve services of other providers on a targeted basis. The state saved approximately $5 million in day care funds while maintaining the quality of day care services, and it did so without major resistance from the provider groups. California has been able with its IPM system to begin automation of its licensing and program quality instruments and linking these data with unit cost and service information on providers. In the development of the program quality instruments, a representative sample of providers from across the state played a critical role in the development and implementation of California's IPM system. These links are providing the basis for a child development, decision support system for the Office of Child Development in California.

Indicator Checklist Improves IPM Systems

Very recently, a number of states (Pennsylvania, West Virginia, Michigan, California, Texas, and New York) have begun experimenting with what has been called an "Indicator Checklist." Simply defined, an indicator checklist is a questionnaire or checklist that contains selected items or indicators from a longer, comprehensive instrument that is used as part of an IPM system. The items on the checklist are those that have been determined to be most effective in discriminating between providers that typically receive high overall scores on the comprehensive instrument or provide a high level of quality care and providers that typically receive low overall scores or provide low level of care (Figure 2).

Because of their value in distinguishing between providers who are in compliance and those that are out of compliance, the items on the in-

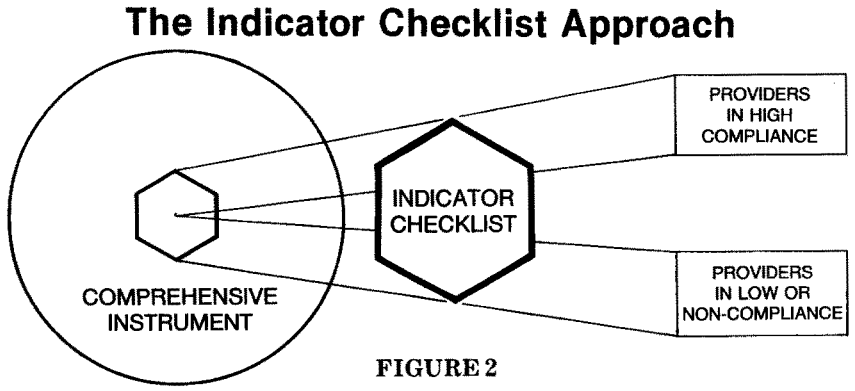


FIGURE 2

dicator checklist have been called "predictor" items. That is, they are a subset of items from the longer instrument that have a strong ability to "predict" the results that would have been obtained had the comprehensive instrument been administered to a given provider. In four of the states mentioned above, the average length of their respective Indicator Checklist's have been approximately 25 items. This compares with the average of approximately 200 items on their respective comprehensive instruments. The relationship between the scores obtained on the state's Indicator Checklists and their comprehensive instruments have been extremely high. When a Pearson's Product Correlation Coefficient was calculated on the Indicator Checklist and the comprehensive instrument for each state the correlation coefficients were always at a r=+.80 or higher (See Figure 2a for a graphic display of West Virginia's data).

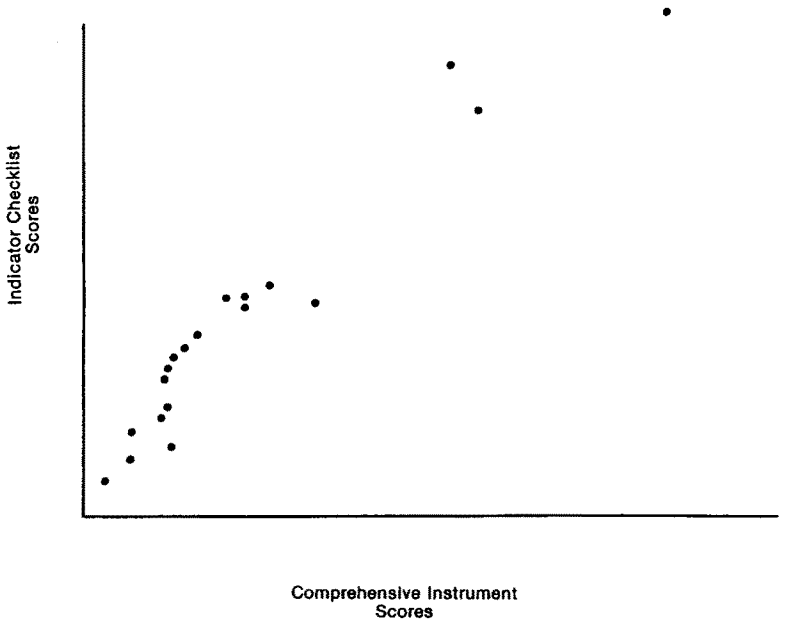


FIGURE 2a

Based on the results of Pennsylvania's, West Virginia's, California's and New York City's Indicator Checklists, certain common items were consistently showing up as predictor items that were separating those good providers from those problem providers. In other words, the following items were always in compliance for the good providers and were always out of compliance for the problem providers:

LICENSING SUBSCALE

A. GROUP SIZE AND ADULT CHILD RATIOS;

INFANTS	1 STAFF TO 5 CHILDREN 10 INFANTS IN A GROUP
TODDLERS	1 STAFF TO 4 CHILDREN 8 TODDLERS IN A GROUP
PRESCHOOLERS	1 STAFF TO 10 CHILDREN 20 PRESCHOOLERS IN A GROUP
SCHOOL AGE	1 STAFF TO 15 CHILDREN 30 SCHOOL AGE CHILDREN IN A GROUP

B. SUFFICIENT SPACE—MINIMUM OF 40 SQ FT PER CHILD;
C. EQUIPMENT IS EASILY ACCESSIBLE TO CHILDREN;
D. ALL VEHICLES ARE EQUIPPED WITH AGE-APPROPRIATE SAFETY CARRIERS;
E. CLEANING MATERIALS ARE INACCESSIBLE TO CHILDREN;
F. EMERGENCY CONTACT INFORMATION IS AVAILABLE FOR ALL CHILDREN;
G. ALL STAFF HAVE HAD PERIODIC HEALTH APPRAISALS;
H. ACTIVITIES PROMOTE:
DEVELOPMENT OF SKILLS
SELF-ESTEEM
POSITIVE SELF-IDENTITY
CHOICE OF ACTIVITIES.
(Fiene, 1984)

To most administrators and policymakers, the advantages of a shorter form will be readily apparent. The short form extends the general advantages of an IPM system in three key ways.

First, it substantially reduces the burden on providers, especially those providers that have a record of high compliance and are judged

suitable for use of the short form—it is proposed that these providers be visited once every three years using the comprehensive instrument. In the intervening years, the indicator checklist should be used.

Second, the indicator checklist approach can further reduce a state's cost of monitoring and permit the more efficient reallocation of staff resources to other activities. A cost effectiveness study conducted in West Virginia utilizing their indicator checklist resulted in a savings of 50% staff time in determining the level of compliance of providers (in dollars, this translated to $800 annually per visit saved (Peat, Marwick, & Mitchell 1983). With such a substantial savings in time, program monitors/evaluators could be freed to act more as consultants in providing technical assistance to providers.

Third, reviews of providers may be consolidated where appropriate. For example, state staff who perform fiscal/contract compliance audits of providers might be trained to administer the indicator checklist during their audit.

The total effect of maintaining a strong compliance monitoring capability that is less of a burden on providers and that achieves greater efficiency with lower cost is a higher quality monitoring system.

What is Needed to Develop an Indicator Checklist?

An indicator checklist is constructed as follows (See Figure 3):

1) Begin with an existing, comprehensive instrument that has a sufficiently large number of items so as to make greater efficiency desirable. The relative importance of each item as reflected in some kind of scoring or weighting system must have been established. Many criteria may be used for weighting the individual items. One criterion that is particularly useful for weighting purposes is the extent to which a particular item is related to health, safety, or developmental risks to children.
2) Your state should have used the comprehensive instrument long enough so that it is considered reliable for monitoring purposes; the instrument should have generated data that can be used to distinguish among providers in substantial compliance and weak or non-compliant providers.
3) With an existing, comprehensive instrument and some historical score information, it is possible to use a simple arithmetical formula (phi coefficient) to select those items from the long questionnaire that are most useful in distinguishing be-

tween good and inadequate programs. These distinguishing or "predictor" items form the basis of the indicator checklist (See Fiene & Nixon, 1983) for a detailed explanation of the formula for developing an indicator checklist).

4) The final step is to include on the short form particular questions or items from the comprehensive instrument that are of critical importance to the health and safety of children. Typically, these are items which, if violated, would be sufficient basis for denying or revoking a license for a day care program. Usually, such items are few in number. They are added to the short form with the predictor items to ensure that children will not be jeopardized by any statistical errors that might occur if only the "predictor" items were used.

From this description of the procedure for developing the shortened instrument, it is clear that the essential prerequisites for such a checklist are: 1. a long, comprehensive instrument in which state administrators have confidence; 2. items on the comprehensive instrument that are weighted to indicate their relative importance; 3. sufficient score data from use of the comprehensive instrument to differentiate among better and worse programs; and 4. state commitment to developing a short form instrument.

Specific Concerns of Administrators and Policymakers

It may be useful to address particular concerns of administrators and policymakers who may be interested in or even actively considering developing a shortened form of their state's monitoring or

Constructing The Indicator Checklist

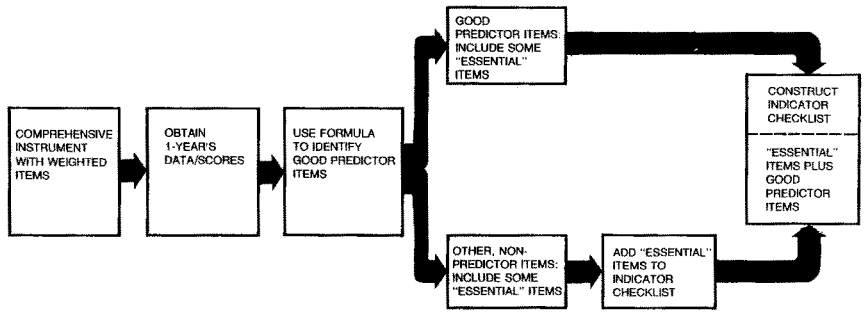

COMPREHENSIVE INSTRUMENT WITH WEIGHTED ITEMS
OBTAIN 1-YEAR'S DATA/SCORES
USE FORMULA TO IDENTIFY GOOD PREDICTOR ITEMS
GOOD PREDICTOR ITEMS: INCLUDE SOME "ESSENTIAL" ITEMS
OTHER, NON-PREDICTOR ITEMS: INCLUDE SOME "ESSENTIAL" ITEMS
ADD "ESSENTIAL" ITEMS TO INDICATOR CHECKLIST
CONSTRUCT INDICATOR CHECKLIST
"ESSENTIAL" ITEMS PLUS GOOD PREDICTOR ITEMS

FIGURE 3

licensing questionnaire or checklist. In particular, administrators will need to know: how their state can make use of an indicator checklist; whether indicator checklists have been tried by other states; how the quality of monitoring can be ensured; and whether there are potential drawbacks.

Can My State Make Use Of An Indicator Checklist?

Practically every state that presently has some form of questionnaire or checklist can potentially profit from using a shortened form of the instrument. Naturally, if your state's instrument is already sufficiently short, then little will be gained by being more selective about questions or items to include. Many states are confronted, however, with lengthy instruments that cover a wide range of requirement areas. These states are prime candidates for short-form instruments.

Similarly, perhaps obviously, if your state does not currently have an instrument-based system, then consideration of an indicator checklist/short form is premature.

In order to develop a successful indicator checklist, it is important that the items on your state's current instrument be clearly linked to:

1. Your state's requirements (regulations); and
2. The results or outcomes that are considered desireable with respect to the providers' performance in such areas as licensing, contract monitoring, and program quality.

Unless there is a clear correspondence between intrument items and requirements, there is a danger that the items selected for inclusion on the short form will be only loosely tied to regulations and may be perceived by providers as improper or illegal. Similarly, if there is only a weak link between items on your state's comprehensive instrument and the results that you expect from providers, then the ground for selecting particular items as good predictors will not be solid enough.

Have Indicator Checklists Been Tried By Other States?

The concept of an indicator checklist may be appealing, but administrators are usually hesitant to take risks that could jeopardize systems that have been developed through years of work. It is often satisfying to know that other states have already tested the concept in practice.

At present, the indicator checklist concept is still an innovation that holds great promise but has been fully implemented in only four

states; Pennsylvania, West Virginia, New York, and California have developed an indicator checklist/short form and are testing the concept. Because the initial analyses conducted by these states suggest that the short form can work, other states such as Michigan and Texas have declared their intention to develop a shortened instrument by using these states' experiences as a guide. Clearly though, the indicator checklist/short-form methodology is still in the experimental stage.

How Can The Quality Of Monitoring Be Ensured?

Top administrators may wonder whether the shortened instrument presented here will compromise the quality of their state's current monitoring effort. Our view is that the short form will enhance current monitoring efforts by increasing the efficient and effective utilization of monitoring staff. But there are precautions that states should take in developing and using indicator checklists.

The indicator checklist/short instrument should not be used as a substitute for the comprehensive instrument, but rather as its complement. If the short form is viewed as the monitoring instrument, then there may be a tendency over time for providers to meet only the requirements covered on the short form. This situation could, indeed, compromise the quality of monitoring.

On the contrary, we would anticipate that states might keep their comprehensive instruments as the definitive set of compliance expectations and administer them for the initial review (e.g., licensing review) of a provider, and could use the indicator checklist/short form as:

1. A screening device to determine whether, for a given provider, it is necessary to administer the longer version; and
2. An interim review instrument to be used as the principal tool for providers who have a good record of compliance.

For example, the comprehensive instrument might continue to be used for "problem" providers and on a periodic basis, say, every three years for good providers. Naturally, if the short form were used with a provider and problems were discovered, then the comprehensive instrument, or some portions of it, could be administered.

Over time, as conditions change, it will be necessary to update and revise both the comprehensive and short instrument. Using the comprehensive instrument at least periodically with all providers will provide a basis for modifying the short form to reflect changing compliance patterns.

We expect that both versions of the instrument would be used by state staff who are trained and competent to assess compliance. These staff would certainly not limit themselves to using the short form if they determined, on site, that conditions warranted using the comprehensive instrument. The purpose of the indicator checklist/short form is to increase the options available to the state for monitoring in a flexible and cost-effective manner, not to put unreasonable constraints or "blinders" on monitoring staff.

What Are The Potential Drawbacks?

As with all innovations, the introduction of an indicator checklist as the basis for routine monitoring in a state may create some problems. Because so few states have introduced indicator checklists on a widespread basis, it is difficult to identify all of the concerns that may arise in practice. However, a few potential problems can be anticipated. (See Table 2).

First, some states' regulations require that all providers be reviewed every year in all regulatory areas. That is, the state insists that a comprehensive review, for example, using the comprehensive form of a state's monitoring instrument, take place for each provider. If this is the situation in your state, then the use of a shortened instrument may depend on changing the current regulatory provisions concerning the frequency and scope of reviews. A strong basis for making such a change is the cost effectiveness of the indicator checklist/short form, that is, its potential for reducing monitoring costs substantially without reducing the quality of the monitoring effort.

TABLE 2

Potential Drawbacks	Possible Solutions
• Regulatory Requirement for Annual Comprehensive Review	• Change Regulatory Requirements
• Staff Resistance	• Educate Staff
• State's Lack of Prerequisites	• Seek Assistance in Obtaining Prerequisites

Second, the state's staff who are responsible for monitoring may resist the introduction of the indicator checklist/short form. From their viewpoint, it may appear that the use of indicator checklists is a reduction in the importance of their professional roles and that the

state's cost savings may take the form of fewer jobs for day care monitors.

In our view, states may need to assure their staff that the indicator checklist/short form is not intended to reduce either the professional judgments involved or the scope of the monitoring function. As mentioned earlier, the comprehensive and short instruments must be used in a complementary way, not as substitutes, in order for the short form to have validity. If anything, the judgment of the monitors may be expanded as it becomes necessary to decide whether, in a particular case, the short instrument will be sufficient to measure compliance with state requirements, and/or program quality criteria. Monitors must be persuaded that the short form is an aid that is designed to reduce the monitors' workload for those providers with whom the short form is appropriate.

The reduction in workload may gradually change the relationship of monitors to providers from one of regulation to one of active support in improving the health and safety of the day care environment and encouraging child development. This change in the monitors' role could enable the state to make even better use of the current monitoring staff's knowledge and experience.

With respect to costs and staff reduction, there is little question that substantial decreases in workload could also result in reduced staffing levels. However, before considering cutbacks in staff, we would encourage states to consider reallocating staff time that is saved because of the short form to other monitoring activities such as technical assistance to providers involving program quality issues.

Third, a state may discover that it does not have the necessary prerequisites, described earlier, to develop and implement an indicator checklist. If your state lacks these prerequisites—in particular a comprehensive instrument, reports of scores, and a system of weighting items on the instrument—then it may be advantageous for you to examine other reports prepared by the Children's Services Monitoring Transfer Consortium that describe how these prerequisites can be met. You may be interested in obtaining the Consortium's series of Guide Books. The three volumes of this series describe in detail how to develop a comprehensive instrument from which an indicator checklist/short form can be derived.

Conclusion

The art of monitoring has evolved considerably in recent years as more highly trained staff have been given responsibility for monitoring, and as clearer procedures, such as instrument-based program monitoring, have been implemented. This evolution has con-

tributed positively to achieving the desirable outcomes of improved day care for children for which the state has developed regulations. At the same time, the evolution has, we hope, made it possible for providers to operate more effectively with the minimum necessary oversight by the state.

Instrument Based Program Monitoring Systems are now being developed in other children's services such as MH/MR services. Pennsylvania has developed its child welfare information system based on the instrument based program monitoring concept. This system meets two needs for Pennsylvania: it tracks children through its foster care system; and it complies with PL 96-272—the Adoption Assistance and Foster Care Act—a federal law. West Virginia is attempting to use the IPM methodology in monitoring its family day care home programs.

Also, a micro-computer, decision support system based on the Instrument Based Program Monitoring and Indicator Checklist methodology is being developed by the Children's Services Monitoring Transfer Consortium (CSMTC). The CSMTC is a group of states (Pennsylvania, West Virginia, California, New York, Michigan, and Texas) who have been disseminating exemplary monitoring techniques from state to state. Based on the combined efforts of these states, a generic indicator checklist that measures compliance with state regulations as well as program quality has been developed (Fiene, 1984). The CSMTC feels that this generic indicator checklist can be used by states who have not developed an instrument to assess providers, or as a model instrument to assist states in developing their own instruments.

The real potential of monitoring in achieving social goals, (such as protecting the health and safety of young children, ensuring quality child development programs, and tying these to child development outcomes), will be better realized through continuing research and development of improved monitoring procedures. It is in this context that the development of the indicator checklist represents a major advance in monitoring children's services.

References

Aronson, S., Fiene, R. & Douglas, E. *Pennsylvania Child Development Program Evaluation Instrument,* Bureau of Child Development, Harrisburg, Pennsylvania, 1977.

Belsky, J.& Steinberg, L. The effects of day care: A critical review, *Child Development,* 1978, *49,* 929-949.

Class, N. & Orton, R. Day care regulation: The limits of licensing, *Young Children,* Vol. #2, 1980, September, 12-17.

Collins, R. Child care and the states: The comparative licensing study, *Young Children,* Vol. #2, 1983, July, 3-11.

Cohen, M. *New York City Indicator Checklist,* NYC Agency for Child Development, New York City, 1983.

CWLA, Day Care Standards Revisions, *Child Welfare League of America,* New York City, 1983.

Douglas, E. & Fiene, R. Making the almost impossible—possible: Evaluation of a human service (Ecological Paradigm for child care monitoring), paper presented at *National Association for the Education of Young Children,* November 1979, Atlanta, Georgia.

Ferrar, H., Gleason, D. & Smith, B. *A State-of-the-Art Report on State Monitoring of Child Care Facilities,* Social Services Research Institute, Washington, D.C., 1980.

Fiene, R., The two year old: Characteristics and management of his play, *Dimensions,* 1974, *2,* 2, 46-58.

Fiene, R., Current trends in research, *Children,* 1975, Spring, 15-18.

Fiene, R. *Child Development Program Evaluation Scale,* Children's Services Monitoring Consortium, Washington, D.C., 1984.

Fiene, R., Douglas, E., & Kroh, K. *Child Development Program Evaluation,* Bureau of Child Development, Harrisburg, Pa., 1977.

Fiene, R., Douglas, E. & Nixon, M. *System Documentation for Pennsylvania's Instrument Based Program Monitoring Information System,* Children's Services Monitoring Services Consortium, Washington, D.C., 1983.

Fiene, R. & Nixon, M. *Instrument Based Program Monitoring System,* Children's Services Monitoring Consortium, Washington, D.C., 1981.

Fiene, R. & Nixon, M. *Indicator Checklist Methodology,* Children's Services Monitoring Consortium, Washington, D.C., 1983.

Harms, T. & Clifford, R. *Early Childhood Environment Rating Scale,* Teachers College, Columbia University, New York, 1980.

HEW, *FIDCR Report of Findings and Recommendations,* U.S. Department of Health Education and Welfare, Washington, D.C., 1978.

Lazar, I. *Lasting Effects After Preschool,* U.S. Department of Health, Education and Welfare, Washington, D.C., 1979.

Merrill, B. *West Virginia Indicator Checklist,* West Virginia Department of Social Services, Charleston, W.V., 1981.

Morgan, G. *Regulation of Early Childhood Programs,* Washington, D.C.: Day Care and Child Development Council of America, Inc., 1971.

Morgan, G. *Alternatives for Regulation of Family Day Care Homes for Children,* Washington, D.C.: Day Care and Child Development Council of America, Inc., 1974.

Morgan, G. Regulating early childhood programs in the eighties, unpublished manuscript, 1980.

NAEYC, Criteria for quality early childhood programs, *National Association for the Education of Young Children,* Washington, D.C., 1983.

Peat, Marwick, & Mitchell, Assessment of the inplementation of the indicator checklist for day care centers in West Virginia, *Children's Services Monitoring Consortium,* September, 1983.

Poole, J. *Child Development Program Quality Review,* California State Department of Education, Sacramento, CA., 1982.

Ruopp, R. *Children at the Center,* Cambridge, MA: Abt Books, 1979.

Watson, S., Fiene, R., & Woods, L. *Kids MicroComputer Software,* Children's Services Monitoring Consortium, Washington, D.C., 1984.

Zigler, E. & Gordon, E. *Day Care: Scientific and Social Policy Issues,* Boston, Massachusetts: Auburn House Publishing Company, 1982.

Stepping Stones (3rd Edition) Key Indicators

Richard Fiene, Ph.D.

April 2013

This short paper will present the Key Indicators as they appear in *Stepping Stones* (3rd edition). It provides the statistically predictive standards (Key Indicators) that could determine overall compliance with *Stepping Stones* (AAP, APHA, NRC, 2013) and *Caring for Our Children* (AAP, APHA, NRC, 2011) based upon the statistical methodology (Fiene & Nixon, 1985). But before delineating the Key Indicators a few definitions need to be provided to put these key indicators in perspective.

Definitions:

Risk Assessment (RA) - a differential monitoring approach that employs using only those rules, standards, or regulations that place children at greatest risk of mortality or morbidity if violations/citations occur with the specific rule, standard, or regulation. *Stepping Stones* (3rd edition) is an example of a risk assessment approach.

Key Indicators (KI) - a differential monitoring approach that employs using only those rules, standards, or regulations that statistically predict overall compliance with all the rules, standards, or regulations. In other words, if a program is 100% in compliance with the Key Indicators the program will also be in substantial to full compliance with all rules, standards, or regulations. The reverse is also true in that if a program is not 100% in compliance with the Key Indicators the program will also have other areas of non-compliance with all the rules, standards, or regulations. The key indicators put forth in this paper are an example of the approach.

Differential Monitoring (DM) - this is a relatively new approach to determining the number of visits made to programs and what rules, standards, or regulations are reviewed during these visits. There are two measurement tools that drive differential monitoring, one is Weighted Risk Assessment tools and the other is Key Indicator checklists. Weighted Risk Assessments determine how often a program will be visited while Key Indicator checklists determine what rules, standards, or regulations will be reviewed in the program. Differential monitoring is a very powerful approach when Risk Assessment is combined with Key Indicators because a program is reviewed by the most critical rules, standards, or regulations and the most predictive rules, standards, or regulations. See Fiene's Logic Model & Algorithm for Differential Monitoring *(DMLMA©)(*Fiene, 2013).

Early Childhood Program Quality Indicator Model (ECPQIM)(Fiene, 2013; Fiene & Kroh, 2000; Griffin & Fiene, 1995; Fiene & Nixon, 1985) – this definition is provided to place the results of this paper into the larger program monitoring systems perspective. ECPQIM are models that employ a key indicator or dashboard approach to program monitoring. Major program monitoring systems in early care and education are integrated conceptually so that the overall early care and education system can be assessed and validated. With these models, it is possible to compare results obtained from licensing systems, quality rating and improvement systems (QRIS), risk assessment systems, key indicator systems, technical assistance, and child development/early learning outcome systems. The various approaches to validation (Zellman & Fiene, 2012) are interposed within this model and the specific

expected correlational thresholds that should be observed amongst the key elements of the model are suggested. **Key Elements** of the model are the following: **CI** = Comprehensive Instrument - state or federal standards, usually rules or regulations that measure health and safety - *Caring for Our Children or Head Start Performance* Standards will be applicable here. Quality Rating and Improvement Systems (QRIS) standards at the state level; *ERS (ECERS, ITERS, FDCRS), CLASS, or CDPES* (Fiene & Nixon, 1985). **RA** = Risk assessment tools/systems in which only the most critical rules/standards are measured. *Stepping Stones* is an example of this approach**.** **KI** = Key indicators in which only predictor rules/standards are measured. The *Thirteen Indicators of Quality Child Care (Fiene, 2003)* is an example of this approach. **DM** = Differential monitoring decision making in which it is determined if a program is in compliance or not and the number of visits/the number of rules/standards are ascertained from a scoring protocol. **T**echnical assistance/training and/or professional development system which provides targeted assistance to the program based upon the Differential Monitoring results. And finally, child outcomes which assesses how well the children are developing which is the ultimate goal of the system.

The Key Indicators from *Stepping Stones* (3rd Edition)[1]

1.1.1.2 - Ratios for Large Family Child Care Homes and Centers

1.3.1.1 - General Qualifications of Directors

1.3.2.2 - Qualifications of Lead Teachers and Teachers

1.4.3.1 - First Aid and CPR Training for Staff

1.4.5.2 - Child Abuse and Neglect Education

2.2.0.1 - Methods of Supervision of Children

3.2.1.4 - Diaper Changing Procedure

3.2.2.2 - Handwashing Procedure

3.4.3.1 - Emergency Procedures

3.4.4.1 - Recognizing and Reporting Suspected Child Abuse, Neglect, and Exploitation

3.6.3.1 - Medication Administration

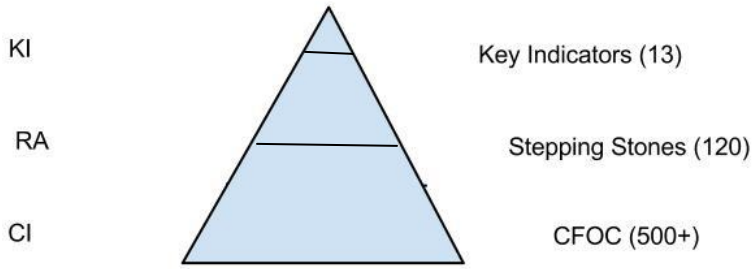
5.2.7.6 - Storage and Disposal of Infectious and Toxic Wastes

6.2.3.1 - Prohibited Surfaces for Placing Climbing Equipment

7.2.0.2 - Unimmunized Children

9.2.4.5 - Emergency and Evacuation Drills/Exercises Policy

Relationship of Key Indicators (KI), Stepping Stones (RA), and Caring for Our Children (CFOC)(CI)



The above diagram depicts the relationship amongst KI, RA, and CI in which the full set of rules is represented by CFOC - Caring for Our Children, followed by RA which are the most critical rules represented by Stepping Stones, and finally the predictive rules represented by the 13 Key Quality Indicators.

Just as there has been three editions of *Caring for Our Children* and *Stepping Stones*, this paper and the resulting Key Indicators represents the third edition of Key Indicators for early care and education. The first two editions are represented in the publications by Fiene & Nixon (1985) and Fiene (2003) respectively (see the reference list below).

References[2]

AAP, APHA, NRC (2011). *Caring for our children: National health and safety performance standards; Guidelines for early care and education programs. 3rd Edition*. Elk Grove Village, IL: American Academy of Pediatrics; Washington, DC: American Public Health Association.

AAP, APHA, NRC (2013). *Stepping stones to caring for our children: National health and safety performance standards; Guidelines for early care and education programs. 3rd Edition*. Elk Grove Village, IL: American Academy of Pediatrics; Washington, DC: American Public Health Association.

Fiene (2013). *Differential monitoring logic model & algorithm*, Research Institute for Key Indicators, Middletown, Pennsylvania.

Fiene (2003). Licensing related indicators of quality child care, *Child Care Bulletin*, Winter 2002-2003, pps 12-13.

Fiene (1985). Measuring the effectiveness of regulations, *New England Journal of Human Services, 5*(2), 38-39.

Fiene (1981). A new tool for day care monitoring introduced by children's consortium, *Evaluation Practice, 1*(2), 10-11.

Fiene & Kroh (2000). Licensing Measurement and Systems, *NARA Licensing Curriculum*. Washington, D.C.: National Association for Regulatory Administration.

Fiene & Nixon (1985). Instrument based program monitoring and the indicator checklist for child care, *Child Care Quarterly, 14*(3), 198-214.

Griffin & Fiene (1995). *A systematic approach to policy planning and quality improvement for child care: A technical manual for state administrators*. Washington, D.C.: National Center for Clinical Infant Programs-Zero to Three.

Zellman & Fiene (2012). *Validation of quality rating and improvement systems for early care and education and School-age care*, Washington, D.C.: OPRE and Child Trends.

Notes:

1 Please see Stepping *Stones (3rd edition)* and *Caring for Our Children (3rd edition)* for the details of each Key Indicator.

2 For the reader who is interested in learning more about the DMLMA/ECPQIM model, please refer to these publications which are available through the following website:

http://RIKInstitute.wikispaces.com

For additional information, please contact:
Richard Fiene, Ph.D., Director
Research Institute for Key Indicators
41 Grandview Drive
Middletown, PA. 17057
DrFiene@gmail.com
RIKI.Institute@gmail.com
717-944-5868 Phone and Fax
http://RIKInstitute.wikispaces.com
http://pennstate.academia.edu/RickFiene

Caring for Our Children Basics: A Brief History of Early Care and Education Standards and their Monitoring in the United States (USA)

Richard Fiene, Ph.D., Senior Research Consultant, rfiene@naralicensing.org

National Association for Regulatory Administration, https://www.naralicensing.org/key-indicators

November 2021

It all started in and around 1965 when the Federal government got into early care and education (ECE) in earnest with Head Start and federally funded day care for low-income families. It started off slowly but began to pick up momentum with exciting studies and research applying principles from developmental psychology to policy making. Researchers and policy makers wanted to make sure that these new programs were not detrimental to young children since our frame of reference were children being raised in orphanages and the ultimate outcome for children was not positive. Would ECE have the same impact?

Issues around quality, appropriateness of standards, and demonstration programs became the focal point of federal research funding. The focal point of this essay is on the appropriateness of the ECE standards and the resulting monitoring systems that were to become key to the federal involvement in early care and education. This essay will be organized by the following 50 years neatly broken out by each decade to get us from this beginning in 1965 until the publication of *Caring for Our Children Basics* in 2015 by the federal government, the Administration for Children and Families, U.S. Department of Health and Human Services. A look at the 2020 decade with a future note is also appended to this essay.

1970s

During the 1970's, the federal government became concerned about what were to be the standards for this new national program related to federally funded ECE for low-income families and their children. Head Start was a separate entity and we will revisit Head Start later but our focus for now is on the federally funded programs which became known back then generically as day care. This nomenclature changed to child care and to finally early care and education (ECE) during this 50-year history. The initial standards for day care were the *Federal Interagency Day Care Requirements (FIDCR)*. A very large appropriateness research study led by Abt Associates to determine what were the most salient standards and their intended impact on children while in day care was conducted during this decade. These standards were to be federally mandated requirements for any program receiving federal funding. This is where group size and adult-child ratios standards became such important safeguards and surrogates for children's health and safety in day care programs.

It also became of interest for the federal government to design the monitoring system that would determine compliance with the FIDCRs. But it became clear to the original designers of this new system that the monitoring of the FIDCR was going to be difficult to do across the full USA. So, the question became, is there a way to monitor the standards in the most effective and efficient manner? This

question and the future of the FIDCR were to be altered and put on hold once we moved into the next decade.

1980s

A change in federal administration and a resulting change in philosophy related to the federal role in America altered many things and one of them was the relationship of the federal government and the states. Rather than the federal government mandating day care requirements, the focus changed with the locus of control moving from the federal level to the state level via block grant funding with very few federal requirements. This meant a moratorium to FIDCR and its ultimate demise. The federal government was not going to be in the business of providing day care, this was going to be the jurisdiction of the states. Head Start did become the exception to this rule with its own standards and monitoring system.

The focus of federal funding switched from the national to the state level in determining compliance with each state's respective child care licensing rules and not with an overarching FIDCR. There was still interest in making these state monitoring systems as effective and efficient but there was no interest in the federal government determining what these requirements would be. Two monitoring approaches grew out of this need for effectiveness and efficiency: risk assessment and key indicators. These two approaches were originally designed and implemented as part of a federally funded project called the Children's Services Monitoring Transfer Consortium in which a group of five states: New York, Michigan, Pennsylvania, West Virginia, and California teamed up to explore their most effective and efficient monitoring systems and begin transferring these systems to one another and beyond.

These two monitoring approaches were tested in the above respective states and it was determined that their impact had a positive effect on the children who were in those day care centers. This was a major finding, similar to the FIDCR appropriateness study, in which these approaches provided safeguards related to the health and safety of children while in day care.

1990s

By the 1990s, it became clear that the federal government had pretty much drawn back from any leadership role in having mandated federal requirements when it came to health and safety in child care. It was left to national ECE advocates who were positioned within the federal government (Administration for Children and Families; Maternal and Child Health Bureau) as well as throughout the USA with national and state agencies and organizations (American Academy of Pediatrics; American Public Health Association, National Resource Center for Health and Safety in Child Care) that saw a need for child care health and safety recommendations at least. If we could not have requirements, we could at least have recommendations and provide guidance to child care programs throughout the USA.

This led to the first edition of *Caring for Our Children* which was a comprehensive set of child care health and safety standards. It was a major game changer for the ECE field because now there was a universal set of standards based upon the latest research literature for states to use as they considered revising and updating their respective state licensing child care rules.

But there was a problem. *Caring for Our Children* was a comprehensive set of health and safety standards which was their strength but at the same time it was their weakness. They were so

comprehensive (well over 500 well researched standards) that they were intimidating and it was difficult to determine where to begin for the states.

Several researchers remembered the two approaches to monitoring designed in the previous decade and wondered if they could be helpful in focusing or targeting which of the standards were the most critical/salient standards. The risk assessment approach to monitoring appeared to have the most immediate applicability and *Stepping Stones* to *Caring for Our Children* was born. This document clearly articulated which of the 500+ *Caring for Our Children* standards placed children at greatest risk for mortality or morbidity by not being in compliance with the respective standard. Since the early 1990s, *Caring for Our Children* and *Stepping Stones to Caring for Our Children* have gone through three editions and have become very important resources to state licensing agencies as they revise, update and improve their ECE rules.

2000s

In this decade several federal and national organizations began to use *Caring for Our Children* standards in innovative ways to measure how well ECE looked at a national level. The Assistant Secretary's Office for Planning and Evaluation in the U.S. Department of Health and Human Services published the *Thirteen Indicators of Quality Child Care* based upon a core set of predictor standards from *Caring for Our Children*. These were standards that predicted overall compliance with all the standards and were seen as an efficient monitoring system. NACCRRA (National Association for Child Care Resource and Referral Agencies) began publishing a national report card on how well states met specific standards and monitoring protocols based upon similar predictor standards from *Caring for Our Children*.

These efforts helped states to make significant changes in their ECE rules in their respective states and in a very voluntary way suggested a means for national standards for the ECE field although we would need to wait until the next decade in order to see such a published document of national ECE health and safety standards for early care and education: *Caring for Our Children Basics*.

2010s

By the 2010s, ECE had grown into a very large but unwieldly assortment of programs with varying levels of quality. Again because of major federal funding, the Child Care Development Block Grant, along with changes and enhancements in professional development, accreditation systems, quality rating and improvement systems, the ECE landscape had become more complex and less easy to navigate. And rather than coming together it was clearly more fragmented than ever.

We had very minimal requirements for the federal funding and most of these requirements were geared to the state agency using the state's respective licensing rules as the threshold for standards. This approach worked well with states with excellent licensing rules, but it wasn't working as well with states who did not have equally excellent licensing rules. We still did not have a core set of standards for ECE programs. Enter *Caring for Our Children Basics* which took the best aspects of the above two monitoring approaches, risk assessment and key indicators and molded it into this new document. This work was led by the federal government's Administration for Children and Families, U.S. Department of Health and Human Services and although the standards are still recommendations and guidance, it is our best attempt at having national standards for early care and education. It is an attempt to provide guidance to the full ECE field, child care, Head Start, preschool, and center based as well as home-based care. It

would be nice to have *Caring for Our Children Basics* as the health and safety foundation for early care and education throughout the USA. I don't see this happening in my lifetime.

2020s: Looking to the Future

As a footnote to this essay, the new decade has been dealt with a major curve ball with COVID19 rearing its ugly head and ECE has been impacted greatly because of this pandemic. As of this writing we are nowhere closer to a solution to getting ECE programs back on line. If anything, the pandemic really demonstrated the fragility of the ECE system we have built over the past 50 years and it clearly has not done very well. My hope is that we can learn from the past 50 years and not continue another 50 years along the same route; although I am guessing that many ECE advocates would be glad to have what we had before the pandemic because what we have right non-sustainable. We know a lot more today than what we knew back in 1965 when we were worried about would day care hurt children's development. We know today that quality ECE benefits children but unfortunately, we are no closer to attaining this today than we were 50 years ago.

Two programs that have been very successful in avoiding these pitfalls are Head Start and the national Military Child Care program. Both programs are exemplary examples of quality early care and education being provided with separate funding streams and standards. Interesting enough when the Administration for Children and Families published *Caring for Our Children Basics*, both these programs were part of the reach of the published standards. As we re-invent and re-structure ECE we should be looking to both these very successful programs for guidance.

Richard Fiene, Ph.D. is a research psychologist and senior research consultant with the National Association for Regulatory Administration and retired professor of psychology, Penn State University. Dr Fiene has spent his professional career in designing and monitoring licensing and quality standards and systems for early care and education.

National Resource Center for Health and Safety in Child Care and Early Education

A Parent's Guide To

Choosing Safe and Healthy Child Care

Revised July 2019

A Parent's Guide To
Choosing Safe and Healthy Child Care

More and more, research tells us that our children's healthy development depends on safe and positive experiences during the first few years of life. If you are a parent/guardian who works during these early years, choosing good child care is one of the most important decisions you will ever make for your child.

To help you make the right choice for your child, researchers have identified 13 guidelines to think about when choosing a child care program. Research shows that if a program follows guidelines, it is more likely to be a safe and healthy place for your child. Considering these guidelines can help you find a place where you can feel comfortable leaving your child.

You might want to visit several different child care programs, either centers or family child care homes, before you decide which one is best for your family. Call each child care program and schedule an appointment for your visit. Once you are there, stay for at least an hour to watch activities, check the surroundings, and ask questions. Place checkmarks on this form to note which guidelines are met in each of the programs you visit. The checklist provides a place to make notes on up to 3 different child care programs.

Child Care Program Names (fill in): *A* *B* *C*	*Child Care Programs* A	B	C
Supervision *Hint: Discipline should be positive, clear, consistent, and fair.*			
✔ Are children supervised at all times, even when they are sleeping?	☐	☐	☐
✔ How do the caregivers/teachers discipline children?	☐	☐	☐
Handwashing and Diapering ***Hint**: Hands should be scrubbed with soap and water for at least 20 seconds and then rinsed and dried. The water faucet should be turned off with a paper towel.*			
✔ Do all caregivers/teachers and children wash their hands often, especially before eating and after using the bathroom or changing diapers?	☐	☐	☐
✔ Is the place where diapers are changed clean?	☐	☐	☐
✔ Do caregivers/teachers always keep a hand on the child while diapering?	☐	☐	☐
✔ Do caregivers/teachers remove the soiled diaper without dirtying any surface not already in contact with stool or urine?	☐	☐	☐
✔ Do caregivers/teachers clean and disinfect the surface after finishing the changing process?	☐	☐	☐

	Child Care Programs A	B	C
Director Qualifications			
✔ Does the director of a child care center have a bachelor's degree in a child-related field?	☐	☐	☐
✔ Has the director worked in child care for at least 3 years?	☐	☐	☐
✔ Does the director understand what children need to grow and learn?	☐	☐	☐
Lead Teacher Qualifications			
✔ Does the lead teacher in a child care center have a bachelor's degree in a child-related field?	☐	☐	☐
✔ Has the teacher worked in child care for at least 1 year?	☐	☐	☐
✔ Does the teacher give children lessons and toys that are right for their ages?	☐	☐	☐
Child: Staff Ratio and Group Size ***Hint**: Your child will get more attention if each caregiver/teacher cares for fewer children.*			
✔ How many children are being cared for in the child care program?	☐	☐	☐
✔ How many caregivers/teachers are there?	☐	☐	☐
Immunizations			
✔ Is your child up-to-date on all of the required immunizations?	☐	☐	☐
✔ Does the child care program have records proving that the other children in care are up-to-date on all their required immunizations?	☐	☐	☐
✔ Are the caregivers/teachers up-to-date on all of the recommended immunizations?	☐	☐	☐
Toxic Substances			
✔ Are toxic substances like cleaning supplies and pest killers kept away from children?	☐	☐	☐
✔ Has the building been checked for dangerous substances like radon, lead, and asbestos?	☐	☐	☐
✔ Is poison control information posted?	☐	☐	☐
Emergency Plan			
✔ Does the child care program have an emergency plan if a child is injured, sick, or lost?	☐	☐	☐
✔ Does the child care program have first aid kits?	☐	☐	☐
✔ Does the child care program have information about who to contact in an emergency?	☐	☐	☐

	Child Care Programs A	B	C
Disaster Drills			
✔ Does the child care program have a plan in case of a disaster like a fire, tornado, flood, blizzard, earthquake or acts of violence?	☐	☐	☐
✔ Does the child care program do practice drills once every month?	☐	☐	☐
Child Abuse			
✔ Can caregivers/teachers be seen by others at all times, so a child is never alone with one caregiver/teacher?	☐	☐	☐
✔ Have all caregivers/teachers gone through a background check?	☐	☐	☐
✔ Have the caregivers/teachers been trained how to prevent child abuse, how to recognize signs of child abuse, and how to report suspected child abuse?	☐	☐	☐
Medications			
✔ Does the child care program keep medication out of reach from children?	☐	☐	☐
✔ Are the caregivers/teachers trained and the medications labeled to make sure the right child gets the right amount of the right medication at the right time?	☐	☐	☐
Staff Training/First Aid ***Hint**: When infants are sleeping, are they on their backs with no pillows, quilts, stuffed toys, or other bedding in the crib with them?*			
✔ Have caregivers/teachers been trained how to keep children healthy and safe from injury and illness?	☐	☐	☐
✔ Do they know how to do first aid and CPR?	☐	☐	☐
✔ Have they been trained to understand and meet the needs of children of different ages?	☐	☐	☐
✔ Are all child care staff, volunteers, and substitutes trained on and implementing infant back sleeping and safe sleep policies to reduce the risk of SIDS (Sudden Infant Death Syndrome, crib death)?	☐	☐	☐
Playgrounds			
✔ Is the playground inspected for safety often?	☐	☐	☐
✔ Is the playground surrounded by a fence?	☐	☐	☐
✔ If there is a sandbox, is it clean?	☐	☐	☐
✔ Is the playground equipment safe, with no sharp edges, and kept in good shape?	☐	☐	☐
✔ Are the soil and playground surfaces checked often for dangerous substances and hazards?	☐	☐	☐
✔ Is equipment the right size and type for the age of children who use it?	☐	☐	☐

For more information on choosing a safe and healthy child care setting, contact your local child care resource and referral agency by checking Child Care Aware, a national consumer education parent hotline and web delivery system at: https://www.childcareaware.org/.

The 13 areas above are based on the 2002 research paper by R. Fiene, *13 Indicators of Quality Child Care: Research Update* available at: http://aspe.hhs.gov/hsp/ccquality-ind02/.

Produced by:
National Resource Center for Health and Safety in Child Care and Early Education
University of Colorado Anschutz Medical Campus, Aurora, CO
https://nrckids.org

Supported by:
Initial development of this *Parent's Guide* was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Cooperative Agreement #U46MC09810.

Revised July 2019

The Saskatchewan Early Care and Education Quality Indicators Tool and Validation: The Last Piece of the Puzzle in Creating a Differential Monitoring Approach

Richard Fiene PhD

Research Institute for Key Indicators/Penn State University/National Association for Regulatory Administration

June 2023

Abstract

This validation study involved 30 programs, 90 classrooms and 180 observations of infant, toddler, and preschool classrooms utilizing the ECERS/ITERS and the SKECPQI instruments. Six trained observers collected the data over a two-month period. The analyses clearly demonstrated that the new SKECPQI instrument is a valid and reliable measure of program quality. PQI #2 clearly showed it predictive power in this study. The SKECPQI and PQI #2 correlated very highly with the ITERS and ECERS. The SKECPQI appears to correlate more highly with regulatory compliance violations than the ECERS or ITERS. The ceiling/plateauing effect is not as evident with the SKECPQI as it is with ECERS/ITERS. The Regulatory Compliance Scale (RCS) is a better sorter for regulatory compliance than the violation data. There is a good deal of internal consistency within the SKECPQI Tool just as it is with the ERSs. The Regulatory Compliance Theory of Diminishing Returns was validated in comparing RCS with ECERS/ITERS. Both the SKECPQI Scale and the Regulatory Compliance Scale are introduced as new improvements to measuring quality and regulatory compliance.

INTRODUCTION

This article will delineate the development, piloting and validating of the Saskatchewan Early Care and Education Quality Key Indicators (SKECPQI) Tool. The purpose of the tool is to assess the overall program quality in centered based childcare programs in the Province of Saskatchewan, Canada. The evolution of the tool resulted from a multi-year effort by the Ministry of Education in the Province of Saskatchewan to build an effective and efficient differential monitoring system.

This effort in building a new differential monitoring system started in 2019 and was completed in 2023. The first component of this restructuring was the Saskatchewan Licensing Key Indicator System (2019). This was followed by the Saskatchewan Risk Assessment Rules (2019). Once these were in place and operational, a validation study was conducted to measure that the two methodologies were operating as they should (2020). A work group was initiated in 2019 and completed its work in 2020 on an Early Care and Education Quality Key Indicator Tool (SKECPQI). The tool was put on hold for 2021 because of

the pandemic and a new Canadian Federal initiative to expand childcare services across the province. The tool initiative began again in 2022. The pilot testing and validation occurred in 2023.

The work and these studies in the Province of Saskatchewan by the Ministry of Education is the first demonstration of a full-blown differential monitoring system involving licensing key indicator rules, risk assessment rules, and quality indicators. Besides the development of each tool, each of these tools have been validated as well. All this work was done as a collaborative effort between the Ministry of Education staff and the National Association for Regulatory Administration (NARA) consultant pool. Presently, Saskatchewan's overall system is the best example of a fully developed differential monitoring system for the early care and education field.

This was a monumental effort involving many individuals at the local, provincial, and national levels and many hours of data collection and analysis. All the reports are available on the NARA Website (https://www.naralicensing.org/key-indicators) and the full data set will be available via Mendeley Data Sources (https://data.mendeley.com/datasets/kzk 6xssx4d/1).

BACKGROUND HISTORY

This study and tool grew out of an interest by Saskatchewan Ministry of Education policy makers to establish a balance between regulatory compliance and program quality in the most effective and efficient manner. The Province of Saskatchewan did not have a QRIS (Quality Rating and Improvement System) in place nor plans on developing one. Generally, when a jurisdiction wants to develop a balance between regulatory compliance and program quality with rules/regulations/standards, QRIS's are generally developed and implemented.

In reviewing the research literature on regulatory science, differential monitoring has been a developing approach used by many other jurisdictions in the human service licensing field, especially in the United States and in several other Canadian Provinces. Based upon this review of the research literature and the work of the National Association for Regulatory Administration (NARA) which has been a long-term promoter of this approach and the resulting methodologies of licensing key indicators, risk assessment rules, and most recently quality indicators, a contract was entered into between the Ministry of Education and NARA.

The tool is the direct result of research into identifying licensing and quality key indicators over a 50-year (1970-2022) research effort in which specific methodologies were developed and the differential monitoring approach was tested and implemented in the 1970's. Since that time, a national database which expanded to an international database of common key indicators from jurisdictions' respective key indicator tools. These key indicators resulted in a very similar tool that Saskatchewan is using. In fact, in 2019 when the Saskatchewan work group was established, they started with that specific tool that had been developed (Fiene, 2019). During the 2019-2020 period, the work group made the tool into a more user-friendly tool for Saskatchewan childcare programs.

The big deal with utilizing the key indicator methodology is its ability to statistically predict as if one administered the full tool in question. Therefore, when one administers the first quality indicator in the Saskatchewan Early Care and Education Quality Indicator tool, it is as if they have administered a licensing based regulatory compliance instrument since the quality of staff is a statistically predictive rule (Fiene, 2002a). The same is true in administering the curriculum quality indicator because it is a

statistically predictive standard when looking at overall program quality (Fiene, 2002b). When it comes to QRIS, having communication between staff and parents and parental involvement is a statistically predictive standard for an overall set of QRIS standards (Fiene, 2014). And finally, when administering the ECERS and ITERS or the CIS quality item indicators these are all statistically predictive items for their respective scales as if you had administered the full scales (Fiene, 2002b).

So, as a state/provincial administrator, I would be interested in focusing my efforts on these indicators which reflect compliance with high quality rules/regulations/standards for early care and education. This would be my starting point. I would make sure that my standards reflected quality teachers with the necessary supports such as coaching/mentoring, an early care and education philosophy based upon an emergent curriculum where children are viewed as competent learners, developmentally appropriate curriculum and child assessments, parental and staff communication and participation, and teacher language based/communicative focus when interacting with children in a give and take manner. All this done within a warm and loving style.

An even more efficient and effective way of using the new program quality tool is to pair it with the National Center for Health and Safety in Child Care's *Parental Guide to Choosing Safe and Healthy Child Care (DHHS: Assistant Secretary's Office for Planning and Evaluation, 2019)*. This is a more aggressive and controversial approach, but it is the most efficient way of conducting monitoring visits in the most abbreviated way. However, as efficiency increases, effectiveness may decrease; so, it is a delicate balancing act. This suggested approach builds off a similar suggestion in which only using *Caring for Our Children: Basics (ACF, 2015)* a DHHS Administration for Children and Families publication would be used as the base for regulatory compliance in the United States.

Differential monitoring grew out of a need for jurisdictions to be more effective and efficient in their oversight and inspection efforts of early care and education programs. This started to occur in the late 1960's and 1970's as many more programs were being established. It was becoming clear that the old one size fits all approach to program monitoring was being overwhelmed by the increasing numbers of programs. Also, from an efficiency standpoint it did not make sense to spend the same amount of time with programs that were performing well as those that really needed additional attention. The birth of differential monitoring occurred which at that time it was called inferential inspections (Fiene & Kroh, 2000). Different terminology, same concept.

Since then, differential monitoring has two basic methodologies that have been used successfully over the years: risk assessment and key indicators. The two methodologies have the same results, shortened or abbreviated reviews but they differ in their approaches. Risk assessment as the name implies identifies specific standards that place clients/children at greatest risk or morbidity or mortality if not complied with. Key indicators are specific standards that statistically predict overall regulatory compliance with all rules. Each has their place in the differential monitoring approach depending on the jurisdictions' emphasis. Most recently, to balance the emphasis on regulatory compliance has been the introduction of quality indicators which are specific standards drawn from quality initiatives, such as professional development, program quality tools, and quality rating & improvement systems.

It is and always has been recommended that these methodologies be used together and not separately. This final study undertaken in the Province of Saskatchewan completes the cycle of doing just that in developing a fully functional differential monitoring system with key licensing and quality indicators as well as risk assessment rules.

THE STUDY DESIGN AND METHOD

The design of this study was to provide a validation study of the use of the Saskatchewan Early Care and Education Quality Key Indicators Tool. A convenience sample was selected in which a good variation of overall quality would be present. There were to be three buckets of quality: High, Middle, and Low. These would be defined via ERS scores. Because this was a validation study it was critical to have sufficient variation in the overall quality of programs to test the sensitivity of the new assessment tool.

The below table (Table 1) provided the guidance to the Saskatchewan Ministry of Education policy staff in determining how to collect the program quality data for the research pilot study related to early childhood quality indicators.

Table 1: Selection Process for Study Programs

Quality	Centers	Classrooms	Ages	Levels	ERS	SKECPQI
High	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3
Middle	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3
Low	10	30	10	Infant	A	1
			10	Toddler	B	2
			10	Preschool	C	3

Notes:

A = ITERS (Infants) (B-1yr)
B = ITERS (Toddlers) (1yr-2yrs)
C = ECERS (Preschoolers) (3+yrs)
1 = SKECPQI/Infant (QI items 1-5, 7, 9-10)
2 = SKECPQI/Toddler or Preschool (QI items 1-5, 7, 9-10) or (QI items 1-6, 8-10)
3 = SKECPQI/Preschool (QI items 1-6, 8-10)
SKECPQI = Saskatchewan Early Childhood Program Quality Indicators tool

A total of 6 trained data collectors were needed, 3 for the ERSs and 3 for the SKECPQI. Each observer collected data from 30 classrooms. A data coordinator was utilized who collected all the data, reviewed the scores from the various tools and sent them to NARA. The data collectors were not aware of which centers are in which group, such as High, Middle, or Low

See the Appendix for the Draft of the SKECPQI tool that was used during data collection.

As said earlier, this study involves the validation of the Saskatchewan Early Childhood Quality Indicators Tool (SKECPQI) and involved the collection of new data utilizing the new tool and collecting Early Childhood Environmental Rating Scale (ECERS/ITERS) data as well. Independent contract staff were trained in the use of the SKECPQI as well as having had training on the ECERS/ITERS and were proficiently reliable on the ECERS/ITERS.

A sample of 30 childcare programs who volunteer to be part of this study was selected with 1/3 identified as high quality, 1/3 identified as medium quality, 1/3 identified as low quality. Each program

had both the SKECPQI and the ECERS/ITERS administered to them utilizing two independent observers. The data from the SKECPQI was compared to the ECERS/ITERS to determine the relationship between the two/three scales. The research hypothesis is that there will be a positive relationship between the two/three scales in which those programs that score high on the SKECPQI will score high on the ECERS/ITERS and those that score low on the SKECPQI will score low on the ECERS/ITERS. The ECERS/ITERS will be used as the reference tool for establishing the validity of the SKECPQI.

A training program and all necessary revisions to policies and procedures was conducted as part of this project by a NARA Consultant on both phase 1 and 2. It will be determined later if the SKECPQI will be administered on an ongoing basis by contracted staff or by Ministry staff. Reporting templates were developed as part of this implementation stage. The implementation stage was evaluated to make certain that all components are in place and working as they should.

Timeline: Phase 1: 6 months; Phase 2: 9 months; Training and Implementation Phase: 12 months, will overlap with phase 1 and 2 and extend beyond both. The total time frame will be 24 months (about 2 years), this will include the final report and final evaluation of the implementation stage

RESULTS

The ECERS and ITERS were used to validate the new Saskatchewan Early Care and Education Quality Indicators Tool (SKECPQI). This is standard procedure when conducting a validation study, a recognized empirically based and accepted standard tool is used in correlational analyses to determine if the new tool is measuring the same dimensions as the standardized tool.

The target tool, the Saskatchewan Early Care and Education Quality Indicators, was to be validated against the ECERS and ITERS to determine if there was a quality relationship between the two tools.

The validation analyses involved detailed correlational analyses between the various scales to determine if a relationship existed and how strong that relationship was. But before delving into this relationship and these analyses, an additional analysis was performed given the sophisticated nature of the Saskatchewan monitoring system. Saskatchewan's Ministry of Education's designed differential monitoring system is by far the most analyzed of all jurisdictions to date, so it was suggested to take advantage of this level of detail and build in an additional series of analyses to further test the regulatory compliance theory of diminishing returns in conducting this study. By doing so, Saskatchewan joins the ranks of the Provinces of Alberta and Ontario, the US States of Georgia and Washington, and the US National Head Start program in conducting studies to either confirm or not this theory of regulatory compliance (please see the NARA website on key indicators which contains all the research reports). The following results delineate the data from that portion of the study.

As part of the data collection in addition to collecting data on the ECERS and ITERS as well as the Saskatchewan Early Childhood Program Quality Indicators scale, a summary sheet containing regulatory compliance data was also obtained on each program. These data contained essential demographic information as well as violations from the last inspection along with a rating of the program which was cross referenced to the regulatory compliance data to generate a Regulatory Compliance Scale. This Regulatory Compliance Scale (RCS) had four levels of regulatory compliance: Full, Substantial, Medium, and Low. This RCS is like the regulatory compliance structure used in the previous studies in the above-mentioned jurisdictions in the US and Canada and has been further developed as a more valid means for

measuring and analyzing regulatory compliance (Fiene, 2022). In the Fiene RCS, the following rubric was used: ***Full = 0 violations; Substantial = 1-3 violations; Medium = 4-9 violations; and Low = 10+ violations.***

The first set of analyses was to determine if a correlation existed between the RCS and the ECERS and ITERS. This was the case with the following results: RCS x ITERS for the infant classrooms = .54; $p < .002$; RCS x ITERS for the toddler classrooms = .42; $p < .03$; and RCS x ECERS for the preschool classrooms = .75; $p < .0001$.

The second level of analyses (ANOVA) was to determine if the RCS levels of Full, Substantial, Medium, and Low demonstrated any significant differences in the ECERS and ITERS. The results were the following: Infant classrooms: Low = 3.07; Medium = 4.89; Substantial = 5.06; Full = 4.69; $F = 11.43$; $p < .0001$. Toddler classrooms: Low = 3.50; Medium = 4.56; Substantial = 4.62; Full = 5.06; $F = 2.27$; $p < .11$. Preschool classrooms: Low = 2.78; Medium = 4.39; Substantial = 4.90; Full = 5.12; $F = 16.27$; $p < .0001$. Apart from the toddler classrooms, both the infant and preschool classrooms support the regulatory compliance theory of diminishing returns ceiling and plateauing effect when it comes to measuring program quality as one moves up the regulatory compliance scale.

Table 2: Regulatory Compliance Scale (RCS) and ECERS/ITERS Scores

RCS	Infant Classrooms	Toddler Classrooms	Preschool Classrooms
Low	3.07	3.50	2.78
Medium	4.89	4.56	4.39
Substantial	5.06	4.62	4.90
Full	4.69	5.06	5.12
Significance	$F = 11.43$; $p < .0001$	$F = 2.27$; $p < .11$ NS	$F = 16.27$; $p < .0001$

ECERS, ITERS for Infant classrooms, ITERS for Toddler classrooms (n = 90):

The ECERS score ranged from 1.41 to 6.00. The ITERS for infant classrooms ranged from 2.16 to 5.77; and the ITERS for toddler classrooms ranged from 2.14 to 5.90. The respective means for the ECERS, ITERS-Infant classrooms, and the ITERS-Toddler classrooms were the following: 4.09, 4.39, 4.39. The means and ranges were all consistent.

The correlations of the infant, toddler and preschool classrooms in each of the 30 facilities were the following: Infant and Toddler classrooms = .65; $p < .0001$; Infant and Preschool classrooms = .74; $p < .0001$; and Toddler and Preschool classrooms = .52; $p < .005$. The classrooms demonstrated a great deal of consistency across the various facilities which one would expect.

SKECPQI for Preschool, Infant, and Toddler Classrooms (n = 90):

The SKECPQI score ranged from 13 to 100. The SKECPQI for infant classrooms ranged from 31 to 91 (Mean=60.10); the SKECPQI for toddler classrooms ranged from 13 to 100 (Mean=55.07); and the SKECPQI for preschool classrooms ranged from 25 to 100 (Mean=57.48).

The correlations of the infant, toddler, and preschool classrooms in each of the 30 facilities were the following: Infant and Toddler classrooms = .74; $p < .0001$; Infant and Preschool classrooms = .85; $p <$

.0001; and Toddler and Preschool classrooms = .75; $p < .0001$. The classrooms demonstrated a great deal of consistency across the various facilities which one would hope to be the case with this type of tool or scale. Based upon these results, the inter-correlations were extremely high and show a great deal of stability and are a reliable measure of quality indicators.

SKECPQI #2 showed a great deal of promise as a standalone quality indicator. SKECPQI#2 correlated significantly with ITERS (.56; $p < .0001$), and ECERS (.61; $p < .0001$) and with the overall SKECPQI scores for infant classrooms (.88; $p < .0001$), toddler classrooms (.81; $p < .0001$), and preschool classrooms (.90; $p < .0001$). This quality indicator dealt with philosophy, curriculum planning and programming. This is not the first time that such an indicator was an excellent predictor. This result has been the case in other program quality studies as well (Fiene, Greenberg, Bergsten, Fegley, Carl, Gibbons, 2002b).

The SKECPQI scale demonstrated a great deal of robustness in the data distribution and a good deal of variation in the data set. These are the characteristics of a new tool that you would hope to find in the scale construction and implementation.

Regulatory Compliance Data for Each of the Programs (n = 30):

The Regulatory Compliance Scale (RCS) distributions were the following: Full = 13%; Substantial = 20%; Medium = 37%; and Low = 27%. Generally regulatory compliance data are more skewed than this distribution but because of the nature of this study, facilities were deliberately selected breaking them up into these categories/levels.

The Regulatory Compliance Scale (RCS) actual regulatory compliance violations played out in the following table, these results for the average number of violations were statistically significant ($F = 3.69$; $p < .03$):

Table 3: Regulatory Compliance Scale by the Number of Violations

RCS	Regulatory Compliance Means	Number of Facilities
Low	4.75	8
Medium	3.90	10
Substantial	1.60	5
Full	0	4

Comparing the ECERS and ITERS with SKECPQI and Regulatory Compliance (RCS) Data:

These are the correlations between RCS and SKECPQI for infants, toddlers, and preschool classrooms. RCS x PQI for the infant classrooms = .58; $p < .001$; RCS x SKECPQI for the toddler classrooms = .51; $p < .005$; and RCS x SKECPQI for the preschool classrooms = .60; $p < .001$. The SKECPQI clearly demonstrates its relationship with regulatory compliance. Also, when the SKECPQI is compared with regulatory compliance violation data, the correlations are higher than those obtained in comparing the ERSs to regulatory compliance violation data. And, in fact, the SKECPQI when compared with the RCS appears not to have a ceiling or plateauing effect. It would appear that the SKECPQI is measuring quality in a different way since this effect does not appear evident in the RCS distributions. This result will need to be confirmed in other studies to make certain this relationship holds up. This is a first for comparing regulatory compliance data with program quality data. In the past, either a ceiling or plateauing effect

was always present when looking at the relationship between regulatory compliance and program quality.

Here are the correlations between SKECPQIs and ERSs for infant, toddler, and preschool classrooms: PQI x ITERS for the infant classrooms = .66; $p < .0001$; PQI x ITERS for the toddler classrooms = .53; $p < .003$; and PQI x ECERS for the preschool classrooms = .66; $p < .0001$. These inter-correlations most suggest that the SKECPQI is a valid tool measuring program quality on a different dimension (quality indicators) than the ERS but measuring quality, nonetheless.

A regression analysis determined that with RCS as the dependent variable, ECERS and regulatory violations were statistically significant at the $p < .0001$ with an $R = .91$. This accounted for practically 75% of the variance in being able to determine regulatory compliance.

DISCUSSION

Last piece of the puzzle in creating a differential monitoring system, that is how this report is being characterized. The Province of Saskatchewan has undertaken all the other methodologies utilized in a differential monitoring approach (Please see the NARA website for these reports, the link is hot linked on the first page of this report). Licensing key indicators and risk assessment rules have been implemented successfully. What remained were the Quality Indicators. This report completes the full cycle of validating these last indicators.

With the completion of this validation study, the Saskatchewan Early Childhood Program Quality Indicators Scale could be adapted by other jurisdictions and utilized as a screener methodology. The reason for suggesting this approach is that all the quality indicators are taken from the Key Indicator Methodology and therefore have predictive value when it comes to determining overall quality (Fiene, 2019a). Also, the indicators are drawn from several early care and education delivery systems and quality initiatives, such as licensing, QRIS, quality scales, accreditation, and professional development.

The other significant finding from this study was the additional confirmation of the regulatory compliance theory of diminishing returns in which the results from this study are consistent with the findings from other studies conducted in Canada and the United States. This continues to be a major finding when it comes to comparing regulatory compliance with program quality and the resulting ceiling and/or plateauing effect related to quality scores. Again, from a public policy viewpoint, this finding has significant implications in how licensing decisions are or should be made.

A very interesting finding which was not expected was the fact that when the SKECPQI scores were compared with the regulatory compliance violation data the usual ceiling/plateauing effect did not emerge as in previous studies when these types of analyses were performed. This result needs further exploration to determine why this occurred. In future studies utilizing the SKECPQI, it will be necessary to do similar analyses with regulatory compliance data to ascertain if this same result occurs. At this point, it is difficult to determine if it is characteristic within the SKECPQI that is producing this result, such as a better balance between regulatory compliance and program quality. Only with further study will we be better able to determine the cause of this different result.

CONCLUSION

This article will be read with a certain amount of skepticism in that it suggests using differential monitoring on a much broader scale; however, this report is like several other validation studies conducted by NARA over the past decade which have now clearly demonstrated the validity of the differential monitoring approach. And because of these validation studies, the differential monitoring approach has been utilized by many jurisdictions and has been cited in the United States Federal Legislation that reauthorized the Child Care and Development Block Grant. In the legislation, it is suggested but not required that states entertain the use of the approach. Based upon the latest childcare licensing data, it appears that many states have attempted to utilize the approach.

This study fits with the other regulatory compliance theory reports from states and provinces that have been completed over the past decade by NARA. As mentioned in the **Results and Discussion Sections**, this study is the most comprehensive of the group since the Province of Saskatchewan developed not only risk rules and key indicator rules for licensing but also quality indicators that could be used within their differential monitoring system. This is the first demonstration of this comprehensive approach.

This study completes what was to be a three-year effort but turned into a five-year effort because of the COVID19 Pandemic. Each component of this overall project is well documented on the NARA Key Indicator website. The three major results of this study: confirmation of the regulatory compliance theory of diminishing returns, the introduction of the regulatory compliance scale and the introduction of the Saskatchewan Early Childhood Program Quality Indicators Tool/Scale are all significant contributions to the licensing research literature, but it is this last contribution that needs further development.

The Saskatchewan Early Childhood Program Quality Indicators Tool/Scale is a new program quality tool that is rather robust in measuring quality using key indicators which are taken from various quality initiative studies conducted over the past several decades. The hope is that it will continue within the early care and education field being validated by other researchers and being used to determine the relative scope of program quality in various early care and education settings. We could see the scale being utilized throughout the United States and Canada. It would be an excellent supplement to either the ERS or CLASS tools. It is a simple, straightforward tool that can be easily trained on and administered. It could provide an interesting supplement for licensing staff when they are doing their licensing reviews. In fact, it is intended to be used in conjunction with licensing key indicators and risk rule tools.

Although this was not reported in the **Results Section**, we think it is vitally important to highlight the significant contributions of the licensing staff and others who helped to develop the groupings and levels of regulatory compliance and quality. It was only because of their level of early childhood expertise and their knowledge of the programs that made the sequencing so effective and impactful as an analytical frame of reference.

One last thought is the introduction of the Regulatory Compliance Scale (RCS) as a more logical and robust rubric when comparing regulatory compliance data with program quality. This thought has been presented elsewhere as a possible improvement within licensing measurement and monitoring systems (Fiene, 2022). The scale has been piloted in the past, but this is the first formal test of it in a specific jurisdiction.

For additional information regarding this research validation study, please contact: Richard Fiene PhD, Research Psychologist & Regulatory Scientist, Research Institute for Key Indicators, Penn State University, rfiene@rikinstitute.com

REFERENCES

ACF, (2015). *Caring for Our Children: Basics*, Department of Health and Human Services, Administration for Children and Families, Washington, DC.

Fiene, (2019a). *The Saskatchewan Key Indicator System: The First Step in Developing a Differential Monitoring Approach*, National Association for Regulatory Administration.

Fiene, (2019b). *The Saskatchewan Centre and Home-Based Weighted Risk Assessment Study*, National Association for Regulatory Administration.

Fiene, (2020). *Validation Research Studies of Key Indicator and Risk Assessment Methodologies in the Province of Saskatchewan*, National Association for Regulatory Administration.

Fiene, (2022). *Regulatory Compliance Scale*, Research Institute for Key Indicators.

Fiene, (2002a). *13 Indicators of Quality Child Care: Research Update*, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

Fiene, (2014a*). Qualistar Rating Key Indicator Study*, National Association for Regulatory Administration.

Lahti, Elicker, Zellman, Fiene, (2014b). Approaches to validating childcare quality rating and improvement systems (QRIS): Results from two states with similar QRIS type designs*, Early Childhood Research Quarterly*, Special Issue.

Fiene, (2019c*). Parental Guide to Choosing Safe and Healthy Child Care*, Office of the Assistant Secretary for Planning and Evaluation, 200 Independence Avenue S.W., Room 450G, Hubert Humphrey Building, Washington, DC 20201.

Fiene, Greenberg, Bergsten, Fegley, Carl, Gibbons, (2002b). *Pennsylvania Early Childhood Program Quality Study*, Harrisburg, Pa: Governor's Office.

Fiene & Kroh, (2000). *Licensing Measurement and Systems,* part of the *NARA Licensing Curriculum*, Virginia: National Association for Regulatory Administration.

Also, check out the following websites for additional Differential Monitoring Reports: https://rikinstitute.com or https://www.naralicensing.org/key-indicators

Appendix

Saskatchewan's Early Learning and Child Care Program Quality Key Indicator Instrument (SKECPQI)

The Saskatchewan Program Quality Work Group[1]

March 2023

INTRODUCTION and BACKGROUND to SKECPQI

Ten Quality Key Indicators (QKI) make up the Saskatchewan's Early Learning and Child Care Program Quality Key Indicator Instrument (SKECPQI). The details about each of the Quality Indicators and data collection instructions in order to obtain the necessary data to determine if a program meets the Key Quality Indicators are delineated below for each quality key indicator. Part 1 - Quality Key Indicators (QKI) 1 – 5 will be collected via record or document review, interviewing individuals, or observation. Part 2 - Quality Key Indicators (QKI) 6 – 10 will be collected via observations in the classrooms throughout the assessment.

These ten quality key indicators were taken from previous studies conducted over the past 40 years by Dr Richard Fiene utilizing the Regulatory Compliance Key Indicator metric (RCKIm) that he developed in the late 1970's. These QKI have held up over time and have now been coupled together into this tool and being pilot tested in the Province of Saskatchewan. The original tool was reviewed by a Provincial Ministry of Education Work Group who met during 2019-2020 and made some revisions to the original tool. All these changes are reflected in this version of the SKECPQI (2023).

---

1) Saskatchewan Program Quality Work Group: Kim Taylor, Derek Pardy, Cindy Jeanes, Tanya Mengel, Samantha Ecarnot, Karen Heinrichs, Michelle Vellenoweth, Kristin Jarvis, and Rick Fiene.

PART 1 – Record/Document Review, Interview, Observation Quality Indicators

INDICATOR 1): Number of ECE III Educators

Assessors will review staff records to determine the number of staff who have these credentials in early childhood education. Record the number of ECEs with the appropriate qualifications and divide them by the total number of ECEs to come up with a percent for the center.

How to Measure:

Go to the ***Staff Information Summary*** form to obtain the data for this item. There are two columns that will do this. Under Certification: *Certification Date and Certification Level* (Highest ECE Level Certified). The certification date should be earlier than the date of the review and the actual level of the certification. In this case, we are interested in the number of (ECEIII's). Record the number of ECEIII working at least 65 hours/month. Then record the number of total teaching staff working at least 65 hours/month below as well. Teaching staff is defined as staff who have a responsibility for working with the children and the programming. Determine the percentage by dividing the total number of staff into the total number of ECEIII Certified teaching staff, ECEIII Certified teaching staff is the numerator, and the total number of teaching staff is the denominator (ECEIII/Total number of teaching staff x 100% = Percent).

Scoring for PQI 1:

The total number of ECEIII Certified teaching staff _____ (1.1)

The total number of teaching staff _____ (1.2)

Total ECEIII teaching staff divided by the total number of teaching staff _____ (%).

Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	***1 = 0 to 25%***	***2= 26 to 50%***	***3 = 51 to 75%***	***4 = 76 to 100%***

INDICATOR 2): Stimulating and Dynamic Environment

The criteria for measuring this are drawn from ***Play and Exploration Guide.*** The program is child centered. Children are viewed as competent learners, and they have the freedom to access classroom materials independently without adult intervention. The children are provided with meaningful choices through activity/learning centers. There is evidence of the children's interests and their projects in the learning environment.

How to Measure:

Below is the checklist of items that should be present to assess if the environment is both stimulating and dynamic for the children. You will want to observe that the following items are occurring in the classroom first. If you do not actually observe it occurring, then check the program plan to find documentation that it normally occurs but you just did not observe today. The checklist items would be found in *Play and Exploration* foundational materials.

Quality Early Learning Environments (Please record all that you observe Y or N):

1. Co-teaching is evident. Y/N _____ (2.1)
2. Children are viewed as competent learners & can access materials independently. Y/N ___ (2.2)
3. Authentic and meaningful materials are used with children. Y/N _____ (2.3)
4. Children are provided with meaningful choices. Y/N _____ (2.4)
5. Children's work, art and photos are displayed respectfully. Y/N _____ (2.5)
6. Family photos are displayed in the early learning program. Y/N _____ (2.6)
7. Documentation of learning is displayed and discusses holistic development. Y/N _____ (2.7)
8. Environment reflects the culture and beliefs of the children, families and staff. Y/N _____ (2.8)
9. Variety of books & other print materials are available throughout the classroom Y/N ____ (2.9)
10. A variety of writing materials are accessible to children most of the time. Y/N _____ (2.10)
11. There is evidence of the children's interests & projects in the classroom. Y/N ___ (2.11)

Scoring for PQI 2:

Total up the number of items where you recorded a "Y" above that you observed (curriculum or in classrooms), divide by 11 x 100% to come up with a percent and record here _____ %. Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	***1 = 0 to 25%***	***2= 26 to 50%***	***3 = 51 to 75%***	***4 = 76 to 100%***

INDICATOR 3): Developmentally Appropriate Curriculum Based on Assessments of Each Child

The key for this quality key indicator is that the program is following an individualized prescribed planning document when it comes to curriculum. It does not mean it is a canned program, in fact, it shouldn't if it is based upon the individual needs of each child's developmental assessment. The assessor will ask to see what is used to guide the curriculum. There should be a written document that clearly delineates the parameters of the philosophy, activities, guidance, and resources needed for the particular curricular approach. There should also be a developmental assessment which is clearly tied to the curriculum. The developmental assessment can be home-grown or a more standardized off-the-shelf type of assessment, the key being its ability to inform the various aspects of the curriculum. The purpose of the assessments is not to compare children but rather to compare the developmental progress of individual children as they experience the activities of the curriculum.

The following key elements should be present when assessing this quality indicator.

- 1) The program practices emergent curriculum, allowing the interests of the children to determine the learning content. The curriculum is informed by individual developmental assessments of each child in the respective classrooms.
- 2) The children and educators are co-learners in the exploration of projects.
- 3) Learning activities of the children are documented, displayed in the learning environment and used to plan further learning activities. This can be assessed developmentally.

How to Measure:

Take a sample of 10 individual children's records and consider the above three elements for EACH record. You should be asking yourself if there is a clear link between an assessment and the developmentally appropriate curriculum so that an individualized learning approach is being undertaken and each child's developmental needs are taken into consideration. These records could be formal, such as portfolios kept for each child or a more informal, anecdotal type of record keeping. The key is that there is a record that can be looked at. It is not adequate if the teacher says they do it from memory – it needs to be written down and documented.

Cross check the child's record to the actual curriculum. Record all the instances (Y's) in which this occurs. All three blocks need to be checked for each record (1-10).

Emergent Curriculum is Practiced (3.1)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 1 +

Children and Educators are Co-learners (3.2)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 2 +

Learning Activities are Documented and Displayed and Used to Plan Future Learning (3.3)

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N	6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N

Key Element 3 +

All three key elements must have a Y to get an overall score of Y. If all three key elements have a Y for that individual record, then record Y in the corresponding block in the overall score.

1 Ys =	2 Ys =	3 Ys =	4 Ys =	5 Ys =	6 Ys =	7 Ys =	8 Ys =	9 Ys =	10 Ys =

= Total of All Three Key Elements (3.4)

Scoring for PQI 3:

The number of positive records (all Ys for all three elements) where there is a crosswalk from developmental assessment to curriculum _____

Percent of positive records (all Ys) (divide the number of positive records by 10 x 100%) _____ %. Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	*1 = 0 to 25%*	*2= 26 to 50%*	*3 = 51 to 75%*	*4 = 76 to 100%*

INDICATOR 4): Opportunities for Staff and Families to Get to Know Each Other

There should be activities both within the center as well as off site where staff and parents have opportunities to meet and greet each other. Communication with family members is documented and enables early childhood providers to assess the need for follow-up. Early childhood providers hold regular office hours when they are available to talk with family members either in person or by phone. Family members are encouraged to lead the conversation and to raise any questions or concerns.

How to Measure:

Look for the following 3 examples in policies developed by the program and determine if they have been carried out with families. It will be necessary to interview staff to complete this indicator if you do not find the three examples in policies:

1. The program provides communication, education, and informational materials & opportunities for families that are delivered in a way that meets their diverse needs. Y/N_____ (4.1)
2. The program communicates with families using different modes of communication, and at least one mode promotes two-way communication. Y/N _____ (4.2)
3. The program demonstrates respect and engages in ongoing two-way communication. The program respects each family's strengths, choices, & goals for their children. Y/N ____ (4.3)

Scoring for PQI 4:

Record the number of Yes's (Y's): _____ (Range: 0 – 3) (Divide by 3 x 100% = _____%). Then based on the percentage, you can find the score of 1-4 as per the chart below.

Circle the Appropriate Level	*1 = 0 to 25%*	*2= 26 to 50%*	*3 = 51 to 75%*	*4 = 76 to 100%*

INDICATOR 5): Families Receive Information on Their Child's Progress Regularly Using a Formal Mechanism

Based upon Indicator #3 above, the information gleaned from the developmental assessments should be the focus of the report or parent conference. Parental feedback about the assessment and how it compares to their experiences at home would be an excellent comparison point. All these interactions should be done in a culturally and linguistically appropriate way representing the parents being served.

How to Measure:

Look for the following four examples in policies developed by the program and determine if they have been carried out with families. Record the number of reports completed or parent conferences over the past year. It will be necessary to interview staff to complete this indicator if you cannot determine from records that the conferences or reports were completed.

NOTE: The examples are mutually exclusive and are not additive; the first example is the highest scored, the third example the least scored. After 1-3 are determined, then do the last example.

- 1) The program does have regularly scheduled (at least 2xs/year) parent conferences in which the children's developmental progress is discussed AND provides the family with a report of their child's developmental progress. Y/N _____ (5.1) (Score 3 points). If "Yes" then go to Number 4. If "No", then go to numbers 2 and 3.
- 2) The program has regularly scheduled (at least 2xs/year) parent conferences in which the children's developmental progress is discussed, but it does not provide a report to the parents on their child's developmental progress. Y/N _____ (5.2) (Score 2 points).
- 3) If the program does not have regularly scheduled (at least 2xs/year) parent conferences, does it provide the family with a report of their child's developmental progress. Y/N _____ (5.3) (Score 1 point). Go to Number 4.
- 4) All these interactions are done in a culturally and linguistically appropriate way representing the parents being served. Y/N _____ (5.4) (Score 1 point)

Scoring for PQI5:

Add up the total points based on the Ys; this will range from "0" to "4". The only way a program can receive a "4", is if a program has regularly scheduled parent conferences at least 2xs/year and provides the family with a report of their child's progress; and it is done in a culturally and linguistically appropriate way.

Record the number of points: _____ (Range: 0 - 4)

Total Score for Part 1 = _____

PART 2 - OBSERVATIONS:

For quality key indicators 6, 7 and 8, it is recommended that the licensing consultant refer to the appropriate Environmental Rating Scale (ERS) tool as a reference tool because these indicators are taken directly from these tools. It is also recommended that these be assessed/observed throughout the assessment and not just during key activity times. Please follow the specific instructions and examples as delineated below and in the appropriate ERS tool: ECERS (Items 12 and 13) or ITERS (Item 12). These specific instructions and examples are provided within this tool for ease of administration and data collection. If there are several preschool aged classrooms randomly select one to do your observations.

INDICATOR 6): Educators Encourage Children to Communicate (Preschool Class)

Assessors will need to observe this item when they do their classroom observations. Initially you can ask educators or the director how children are encouraged to communicate but in order to gather reliable and valid information regarding this question/standard, it needs to be observed in the various interactions between staff and children. Things to look for would be more back and forth conversations rather than one-way conversations where educators are telling children what to do. Look for opportunities where children can describe what they are doing, how they feel about what they are doing, and why they are doing particular activities. Educators expand upon children's conversations.

These opportunities can occur anywhere in the classroom or outside, such as in dramatic play, tabletop activities or on the playground. Materials should be present that encourage communication such as toy telephones, puppets, flannel boards, dolls and dramatic play props, small barns, fire stations, or dollhouses. These create a lot of conversation among children as they assume many different roles. Children also talk when there is an interested person who listens to them. The staff in a high-quality early childhood classroom will use both activities and materials to encourage growth in communication skills.

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;
Score the classroom a 1 if the following occur:

- No activities used by staff with children to encourage them to communicate, for example: nontalking about drawings, dictating stories, sharing ideas at circle time, finger plays, singing songs. Y/N _____ (6.1)
- Very few materials accessible that encourage children to communicate. Y/N _____ (6.2)

Score the classroom a 2 if the following occur (If the classroom does not have all 3 indicators but has 2 of the indicators then score this item 1+):

- Some activities are used by staff w/children to encourage them to communicate. Y/N _____ (6.3)
- Some materials are accessible to encourage children to communicate. Y/N ____ (6.4)
- Communication activities are generally appropriate for the children in the group. Y/N _____ (6.5)

Score the classroom a 3 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 2+):

- Communication activities take place during both free play and group times, for example: child dictates story about painting; small group discusses trip to store. Y/N _____ (6.6)
- Materials that encourage children to communicate are accessible in a variety of interest centers, for example: small figures and animals in block area; puppets and flannel board pieces in book area; toys for dramatic play outdoors or indoors. Y/N _____ (6.7)

Score the classroom a 4 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff balance listening and talking appropriately for age and abilities of children during communication activities, for example: leave time for children to respond; verbalize for child with limited communication skills. Y/N _____ (6.9)
- Staff link children's spoken communication with written language, for example: write down what children dictate & read it back to them; help them write notes to parents. Y/N _____ (6.10)

Scoring for PQI 6:
Total up the number of "Y's" and record the appropriate level. In order for a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+". If there is a "+" please also mark it in the box.

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 7): Infant Toddler Observation (if applicable) **(Infant Classroom)**

NOTE: If there is an infant, toddler or combined infant/toddler classroom that needs to be assessed, then use the following ITERS item directly from the ITERS Tool (Item 12), if there is not an infant toddler classroom, then skip to Indicator 8.

Conversations and questions should be used with all children, even young infants. Conversations using verbal and nonverbal turn-taking should be considered when scoring. Most conversations and questions initiated by infants will be nonverbal, such as widening of baby's eyes or waving arms and legs. Observe staff response to such nonverbal communication. For infants and toddlers, the responsibility for starting most conversations and asking questions belongs to the staff. As children become more able to initiate communication, staff should modify their approach in order to allow children to take on a greater role in initiating conversations and asking questions. Staff should provide answers to questions used by children if children cannot answer, and as children become more able to respond, questions should start to include those that the child can answer. If there was not an infant classroom, skip this Indicator and please note that here and on the summary score sheet by marking N/A: _____

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;

Score the classroom a 1 if the following occurs:

- Staff never initiate turn-taking conversations with children, for example: rarely encourage baby to babble back; simple back and forth exchanges with verbal children never observed. Y/N _____ (7.1)
- Staff questions are often not appropriate for children, or no questions are asked, for example: too difficult to answer; carry a negative message. Y/N _____ (7.2)
- Staff respond negatively when children can't answer questions, for example: "You should know this"; "You did not listen". Y/N _____ (7.3)

Score the classroom a 2 if the following occurs (If the classroom does not have all 3 indicators but has 2 of the indicators then score this item 1+):

- Staff sometimes initiate conversations with children, for example: babble back and forth with baby; copy baby's sounds; respond to baby's crying with verbal response; have short back and forth toddler interactions. Y/N _____ (7.4)
- Staff sometimes ask children appropriate questions and wait for the child to respond, for example: ask baby if she likes toy and pay attention as baby smiles; ask toddler what he is eating and wait for him to think of word. Y/N _____ (7.5)
- Staff respond neutrally or positively to children who can't answer questions. Questions asked are sometimes meaningful to children, for example: child responds with interest; does not ignore staff questions. Y/N _____ (7.6)

Score the classroom a 3 if the following occurs (If the classroom does not have all 4 indicators but has 2 or more of the indicators then score this item 2+):

- Staff initiate engaging conversations with children throughout the observation, for example: show enthusiasm; use tone that attracts child's attention. Y/N _____ (7.7)
- Staff often personalize questions and/or conversations for individual children, for example: talk about children's families, preferences, interests; what they are playing with; what they did over weekend; child's mood; use child's name. Y/N _____ (7.8)

- Staff often pay attention to children's questions, verbal or nonverbal, and answer in a satisfying manner for the child. Y/N _____ (7.9)
- Staff ask questions in which children show interest in answering, for example: make the questions funny or mysterious; use attractive tone; meaningful and not too difficult to answer. Y/N _____ (7.10)

Score the classroom a 4 if the following occurs (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff frequently have turn taking conversations with children throughout the observations. Many appropriate questions are used throughout the observation, during both play and routines. Y/N _____ (7.11)
- Staff ask children appropriate questions, wait a reasonable time for child response, and then answer if needed, for example: "Are you hungry? . . . Yes, you are!"; "Where's the ball? . . . These it is! You found the ball". Y/N _____ (7.12)

Scoring for PQI 7:

Total up the number of "Y's" and record the appropriate level. For a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+".

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 8): Educators Use Language to Develop Reasoning Skills (Preschool)

Assessors will need to observe very carefully as this standard can be difficult to determine because it is tying language and cognition together. Again, this opportunity can occur in any setting in or out of the classroom because it is the basis for problem solving through the use of language. Also look for educators redirecting children's conversations when appropriate. Staff should use language to talk about logical relationships using materials that stimulate reasoning. Through the use of materials, staff can demonstrate concepts such as same/different, classifying, sequencing, one-to-one correspondence, spatial relationships, and cause and effect.

How to Measure:

Observe the classroom for a minimum of 15 minutes. Once completed, consider where the classroom falls based on the following scale;

Score the classroom a 1 if the following occur:

- Staff do not talk with children about logical relationships, for example: ignore children's questions and curiosity about why things happen, do not call attention to sequence of daily events, differences and similarity in number, size, shape, cause and effect. Y/N _____ (8.1)
- Concepts are introduced inappropriately, for example: concepts too difficult for age and abilities of children, inappropriate teaching methods used such as worksheets without any concrete experiences; teacher gives answers w/o helping children to figure things out. Y/N _____ (8.2)

Score the classroom a 2 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 1+):

- Staff sometimes talk about logical relationships or concepts, e.g.: explain that outside time comes after snacks, point out differences in sizes of blocks children use. Y/N _____ (8.3)

- Some concepts are introduced appropriately for ages and abilities of children in group, using words and experiences, for example: guide children with questions and words to sort big and little blocks or to figure out why ice melts. Y/N _____ (8.4)

Score the classroom a 3 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 2+):

- Staff talk about logical relationships while children play with materials that stimulate reasoning, for example: sequence cards, same/different games, size and shape toys, sorting games, numbers and math games. Y/N _____ (8.5)
- Children are encouraged to talk through or explain their reasoning when solving problems, for example: why they sorted objects into different groups, in what way two pictures are the same or different. Y/N _____ (8.6)

Score the classroom a 4 if the following occur (If the classroom does not have both indicators but has one of the indicators then score this item 3+):

- Staff encourage children to reason throughout the day, using actual events and experiences as a basis for concept development, e.g.: children learn sequence by talking about their experiences in the daily routine or recalling the sequence of a cooking project. Y/N _____ (8.7)
- Concepts are introduced based upon children's interests or needs to solve problems, for example: talk children through balancing a tall block building, help children figure out how many spoons are needed to set a table. Y/N _____ (8.8)

Scoring for PQI 8:

Total up the number of "Y's" and record the appropriate level. In order for a classroom to receive a particular score, all "Y's" must be checked for the appropriate level (1 - 4) from above or partial credit given in order to obtain a "+".

Circle the Appropriate Level	***1***	***2***	***3***	***4***

For quality key indicators 9 and 10 it is recommended that these be assessed/observed throughout the observation period and not just during key activity times. These two quality key indicators should be observed in two-minute blocks over ten sequences for a total of 20 minutes. These two items should also be used with each age group you are assessing.

INDICATOR 9): Educators Listen Attentively When Children Speak

This quality indicator focuses on the early childhood educator(s) looking directly at the children with nods, rephrases their comments, engages in conversations. Children should have the undivided attention of the specific educator they are addressing. Educators should not be looking away or pre-occupied with others. They should be at the child's level making eye contact. The intent is to observe all children and educators in the room.

How to Measure:

Do this in timed 2-minute observations recording each time you observe this occurring. Record at least 10 different observation periods. These do not need to be consecutive in order to fully observe classrooms and educators. Please use the following scale to assess your recordings: Likert Scale (1-4) where 1 = Never/Not at All; 2 = Somewhat/Few Instances; 3 = Quite a Bit/Many Instances; 4 = Very Much/Consistently):

Make the actual recordings using the Likert Scale (1-4) above for each individual observation and record in each cell below.

10 Observations:

10.1	2	3	4	5	6	7	8	9	10.10

Scoring for PQI 9:

Once all the observations are made, add up the results from the Likert Scale (1-4) and record the total number here: _____ (Range: 10 - 40)(Divide this result by 10) = _____ (1-4)(Round upward or downward to the whole number (3.7 = 4; 2.2 = 2)).

Circle the Appropriate Level	***1***	***2***	***3***	***4***

INDICATOR 10): Educators Speak Warmly to Children

This quality indicator focuses on the early childhood educator(s) always engaging in a caring voice and body language with every child. Educators do not use harsh language or commands in speaking to children, but rather again are on the child's level making eye contact. Think of the way Fred Rogers would engage his audience where you always felt you were the most important person in the world when he talked to the TV.

How to Measure:
Do this in timed 2-minute observations recording each time you observe this occurring. Record at least 10 different observation periods. Please use the following scale to make your recordings: (This item is on a Likert Scale (1-4) where 1 = Never/Not at All; 2 = Somewhat/Few Instances; 3 = Quite a Bit/Many Instances; 4 = Very Much/Consistently):
Make the actual recordings using the Likert Scale (1-4) above for each individual observation and record in each cell below.
10 Observations:

10.1	2	3	4	5	6	7	8	9	10.10

Scoring for PQI 10:

Once all the observations are made, add up the results from the Likert Scale (1-4) and record the total number here: _____ (Range: 10 - 40) (Divide this result by 10) = _____ (1-4).
(Round upward or downward to the whole number (3.7 = 4; 2.2 = 2)).

Circle the Appropriate Level	***1***	***2***	***3***	***4***

SKECPQI Scoring Protocol

LEVEL	Standardized Scores	Actual Scores
High Quality	Mixed Age: 36+ Preschool: 32+ Infant-Toddler: 28+	Mixed Age: _____ Preschool: _____ Infant-Toddler: _____
High - Mid Quality	Mixed Age: 30 – 35 Preschool: 26 - 31 Infant-Toddler: 22 - 27	Mixed Age: _____ Preschool:_____ Infant-Toddler:_____
Mid – Low Quality	Mixed Age: 20 – 29 Preschool: 16 - 25 Infant-Toddler: 12 - 21	Mixed Age: _____ Preschool: _____ Infant-Toddler: _____
Low Quality	Mixed Ages: 19 or less Preschool: 15 or less Infant-Toddler: 11 or less	Mixed Age: _____ Preschool:_____ Infant-Toddler: _____

Note:

Members of the Original Saskatchewan Program Quality Work Group are the following:
Ministry of Education: Kim Taylor, Derek Pardy, Cindy Jeanes, Tanya Mengel, Samantha Ecarnot, Karen Heinrichs, Michelle Vellenoweth, Kristin Jarvis, and NARA Consultant: Rick Fiene.

Additional Information regarding the psychometrics of the tool contact: Richard Fiene, Ph.D., Research Psychologist, Research Institute for Key Indicators & Penn State University. RFiene@RIKInstitute.com or RFiene@NARALicensing.org

10/2020; 4/2021; 1/2023; 2/2023; 3/2023 versions

After completing your observations, reviewing all documentation, and interviewing staff, when necessary, please transfer all your results to the Summary Table below. If there was not an infant classroom, please note here, no infant classroom: _____. If there was not a toddler classroom, please note here, no toddler classroom: _____. If there was not a preschool classroom, please note here, no preschool classroom: _____.

Key Q Indicator	***Quality Indicator Content***	***Scale Source***	***Potential Score***	***Actual Score***
QKI 1	Professional Development	NAEYC	1-4	1, 2, 3, 4
QKI 2	The Environment	Saskatchewan	1-4	1, 2, 3, 4
QKI 3	Curriculum and Assessment	NAEYC	1-4	1, 2, 3, 4
QKI 4	Family Engagement I	QRIS	1-4	1, 2, 3, 4
QKI 5	Family Engagement II	QRIS	1-4	1, 2, 3, 4
QKI 6	Communication (Preschool)	ECERS	1-4 or NA	1, 2, 3, 4, +, NA
QKI 7	*Infant Classroom*	*ITERS*	*1-4 or NA*	1, 2, 3, 4, +, NA
QKI 8	Reasoning Skills (Preschool)	ECERS	1-4 or NA	1, 2, 3, 4, +, NA
QKI 9	Listen Attentively	CIS	1-4	1, 2, 3, 4
QKI 10	Speak Warmly	CIS	1-4	1, 2, 3, 4

Notes:
Use ITERS if: (Infants) (B-1yr)
Use ITERS if: (Toddlers) (1yr-2yr)
Use ECERS if: (Preschoolers) (3yr+)

SKECPQI/Infant (administer QKI items 1-5, 7, 9-10) (Scores 8-32)

SKECPQI/Toddler or Preschool (administer QKI items 1-5, 7, 9-10) (Scores 8-32) or (administer QKI items 1-6, 8-10) (Scores 9-36). Mixed age group (administer QKI items 1-10) (Scores 10-40)

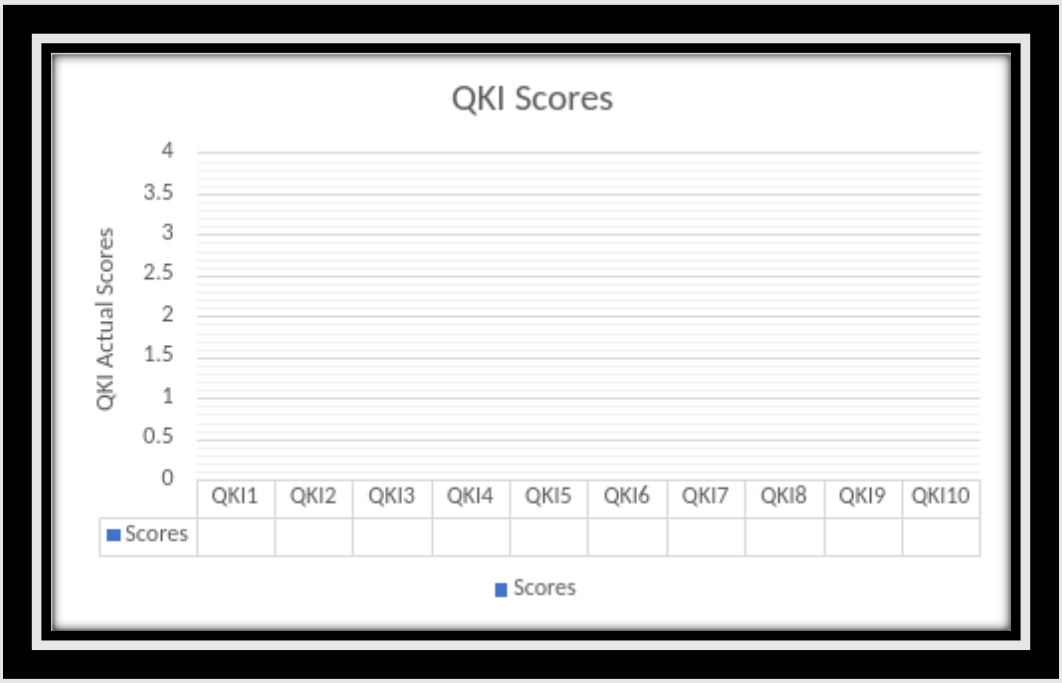
SKECPQI/Preschool (administer QKI items 1-6, 8-10) (Scores 9-36)

All the above 10 quality indicators (SKECPQI) have been taken from other sources having been identified in Quality Indicator Studies conducted by Dr Richard Fiene from 1980 – 2020. Please refer to the source documents for details on their creation: *ECERS, ITERS, QRIS/INQUIRE, CIS/Arnett, NAEYC, SASKATCHEWAN PLAY & EXPLORATION.* For additional information, reports, and publications related to these studies, please go to https://www.naralicensing.org/key-indicators Or https://rikinstitute.com/publications/

SKECPQI: SASKATCHEWAN EARLY CHILDHOOD PROGRAM QUALITY INDICATORS

CHART/GRAPH

	Scores
QKI1	
QKI2	
QKI3	
QKI4	
QKI5	
QKI6	
QKI7	
QKI8	
QKI9	
QKI10	
TOTAL	



QKI and key elements/sub items and comments Scoresheet:

QKI1 _____ 1.1 _____ 1.2 _____ Comments: _____

QKI2 _____ %

2.1 _____ Comments: _____

2.2 _____ Comments: _____

2.3 _____ Comments: _____

2.4 _____ Comments: _____

2.5 _____ Comments: _____

2.6 _____ Comments: _____

2.7 _____ Comments: _____

2.8 _____ Comments: _____

2.9 _____ Comments: _____

2.10 _____ Comments: _____

2.11 _____ Comments: _____

QKI3 _____ %

3.1 _____ Comments: _____

3.2 _____ Comments: _____

3.3 _____ Comments: _____

3.4 _____ Comments: _____

QKI4 _____ %

4.1 _____ Comments: _____

4.2 _____ Comments: _____

4.3 _____ Comments: _____

QKI5 _____ Points

5.1 _____ Comments: _____

5.2 _____ Comments: _____

5.3 _____ Comments: _____

5.4 _____ Comments: _____

QKI6 _____ Level

6.1 _____ Comments: _____

6.2 _____ Comments: _____

6.3 _____ Comments: _____

6.4 _____ Comments: _____

6.5 _____ Comments: _____

6.6 _____ Comments: _____

6.7 _____ Comments: _____

6.8 _____ Comments: _____

6.9 _____ Comments: _____

QKI7 _____ Level

7.1 _____ Comments: _____

7.2 _____ Comments: _____

7.3 _____ Comments: _____

7.4 _____ Comments: _____

7.5 _____ Comments: _____

7.6 _____ Comments: _____

7.7 _____ Comments: _____

7.8 _____ Comments: _____

7.9 _____ Comments: _____

7.10 _____ Comments: _____

7.11 _____ Comments: _____

7.12 _____ Comments: _____

QKI 8 _____ Level

8.1 _____ Comments: _____

8.2 _____ Comments: _____

8.3 _____ Comments: _____

8.4 _____ Comments: _____

8.5 _____ Comments: _____

8.6 _____ Comments: _____

8.7 _____ Comments: _____

8.8 _____ Comments: _____

QKI9 _____ Level

9.1 _____ Comments: _____

9.2 _____ Comments: _____

9.3 _____ Comments: _____

9.4 _____ Comments: _____

9.5 _____ Comments: _____

9.6 _____ Comments: _____

9.7 _____ Comments: _____

9.8 _____ Comments: _____

9.9 _____ Comments: _____

9.10 _____ Comments: _____

QKI10 _____ Level

10.1 _____ Comments: _____

10.2 _____ Comments: _____

10.3 _____ Comments: _____

10.4 _____ Comments: _____

10.5 _____ Comments: _____

10.6 _____ Comments: _____

10.7 _____ Comments: _____

10.8 _____ Comments: _____

10.9 _____ Comments: _____

10.10 _____ Comments: _____

Quality Key Indicators (QKI)	Elements/Items	Data Collection
1	1.	Record Review
2	11	Policy, Records, Interviews
3	4	Policy, Records, Interviews
4	3 . . .	Policy, Records, Interviews
5	4	Policy, Records, Interviews
6	9	Observation
7	12	Observation
8	8	Observation
9	10	Observation
10	10	Observation
TOTAL	**Potential Score = 78**	**Actual Score Obtained = _____**

Improving Child Care Quality Through an Infant Caregiver Mentoring Project

Richard Fiene
The Pennsylvania State University

ABSTRACT: An evaluation of a mentoring training program for infant caregivers is described. Fifty-two infant caregivers from 27 childcare center-based programs were involved in a four month long intervention in which they were paired with an experienced early childhood educator. The focus of the mentoring program was to improve the overall quality of the classroom environment, as well as making the caregivers more sensitive to the needs of the infants. The results clearly indicated that the mentoring program was very effective in improving the overall quality of the classroom, as well as making caregivers more sensitive to infants' needs.

KEY WORDS: infant caregivers; childcare; mentoring; training.

Introduction

This paper describes a child care mentoring project designed to improve the quality of infant and toddler child care programs in south central Pennsylvania. The goal of the mentoring project was to improve the quality of the child care environment and specifically the quality of caregiver-child interactions. As most caregivers in Pennsylvania only receive workshop training, the goal of this project was to compare the mentoring approach to the more typical workshop training. Mentoring is being explored because of its targeted intensive one-on-one nature in delivering training to caregivers based upon needs assessments. The project was conducted during the later half of 2000 and the beginning of 2001. The results presented in this paper are part of the pre- and post-test data collection phase (summer 2000 and winter 2000–2001) of this mentoring project. The actual mentoring intervention occurred from September through December 2000.

Correspondence should be directed to Richard Fiene, Ph.D., Director, Capital Area Early Childhood Training Institute, and Senior Research Associate, Prevention Research Center, The Pennsylvania State University, 2001 North Front Street, Suite 314, Harrisburg, PA 17102; e-mail: rjf8@psu.edu, Web page: http://caecti.org.

Mentoring in childcare has been documented in the literature for the past 10–15 years (Breunig & Bellm, 1996; Fenichel, 1992). It has been demonstrated to be an effective mode of training/technical assistance (Breunig & Bellm, 1996). However, in the majority of studies conducted there are few, if any, demonstrations that utilize a randomized trial design (Breunig & Bellm, 1996). Many studies track the progress of the intervention group, some studies have comparison groups, but few, if any, have employed a randomized design. This research paper will describe the pre- and post-test data collected as part of a study that has employed a randomized design.

The majority of research (Clarke-Stewart, 1987; Goelman & Pence, 1987; Howes, 1987; Phillips, 1987; Kontos & Fiene, 1987; Galinsky, Howes, Kontos, & Shinn, 1994; Scarr, Eisenberg, & Deater-Deckard, 1994; Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997; Helburn, 1995; Fiene, 1995, 1996; Jorde-Bloom, 1988; Love, Schochet & Meckstroth, 1986) completed on early childhood quality has focused on preschool programs, with infant toddler programs rarely as the central focus of the research. The research completed in infant toddler programs has clearly documented the mediocre level of care provided to children in these programs (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997). In the present study, we focus on the first three years of life. All the centers and the classrooms reported upon in this study serve children from birth to less than three years of age.

This report is organized as follows: a methodology section briefly describes the sample selected with basic demographic information on directors, caregivers and the programs. This is followed by a results section that provides pre- and post-test average scores for each of the assessment tools utilized in this study to measure quality, caregiver behaviors, knowledge, and organizational climate of programs. This section is followed with a discussion section and implications regarding this mentoring project.

Methods

Study Design

This study involved 52 caregivers from 27 sites in south central Pennsylvania. All programs were child care centers licensed by the Department of Public Welfare. Seven of the sites were accredited by the National Association for the Education of Young Children.

This study employed a randomized design in which a self-selected group of programs and caregivers were randomly assigned to two groups, either the mentoring group or the comparison non-mentoring

comparison/control group. Intervention model mentoring group received intensive mentoring from a seasoned early childhood professional (minimum of 5–7 years of experience in the early childhood field as both a director and teacher) from September to December 2000. The mentoring model consisted of a problem solving approach in which the mentor spent a good deal of time observing in the beginning weeks in order to develop a trusting relationship with the protégé. Once both the mentor and protégé felt comfortable then suggestions could be entertained by the mentor.

The comparison group did not receive the mentoring intervention and only had the regular workshop type variety training available to them. However, the comparison group did receive mentoring during the Spring 2001 from March to June 2001. What is of interest in this study is to determine how much the two groups have improved from the pre-test data collection because they were essentially equivalent at that point on all measures.

Programs were recruited by the Capital Area Early Childhood Training Institute, a broad based community focused training institute. Program directors were invited to attend a meeting describing the mentoring project. Of those attending, 95% agreed to participate in the project. Fifty two caregivers started the project, 14 caregivers dropped out of the project between pre- and post-test. There was an equal drop out rate from both the mentoring and the control groups.

Data from the four quality measures used for all the programs are presented in Table 1. The four measures of quality were the Infant Toddler Environment Rating Scale (ITERS), the Arnett Caregiver Observation Scale, the Knowledge of Infant Development (KIDI), and the Bloom Scales of Organization Climate.

The program directors' average age is 31 with a range from 24–53

Table 1
ITERS, Arnett, KIDI, Bloom Scale Scores

All Programs (n = 38)	Pre-Test	Post-Test	Change	Significance
ITERS	134	140	+6	ns
Arnett	30	40	+10	ns
KIDI	14	14	-0-	ns
Bloom	78	79	+1	ns

years of age. They are predominantly Caucasian (81%). Eight percent have associate degrees, 78% have bachelor's degrees, and 14% have master's degrees. They had been employed as directors in their program for an average of 31 months with a range from 1 month to 120 months. Their average pay is between $20000–25000 per year. Sixty percent have health insurance and 45% have some form of dental or life insurance. Forty-five percent are in a retirement system.

The average age of caregivers in the programs was 36 with a range from 18–68. They are predominantly Caucasian (77%). Fifty-seven percent have high school diplomas, 16% have some college credits, 5% have CDA's, 16% have associate degrees, 5% have bachelor's degrees, and 2% have master's degrees. They have been employed as caregivers in their program for an average of 34 months with a range from 1 month to 153 months. They have worked in the early childhood field as caregivers for an average of 71 months with a range from 1 month to 312 months. Their average pay is between $10000–15000 per year. Fifty percent have health insurance and 33% have some form of dental or life insurance. Thirty-three percent are in a retirement system.

The average size of the centers is 98 children with 17 staff employed either full time or part time at the program. The average weekly fee for infant care is $137.00 per week and for toddler care is $124.00 per week. The majority of staff are employed at the centers for either less than 1 year or greater than 5 years.

Results

Both the mentoring and comparison groups were tested for equivalence at the beginning of the project in the pre-test data collection phase. There were no statistically significant differences on any of these measures at the pre-test. When the programs and caregivers were measured at the post-test, positive changes occurred although none were found to be statistically significant. In the aggregate, the programs that continued with the mentoring project showed improvements in the overall quality of care.

Tables 2 through 5 present the pre- and post-test data for the intervention and control groups.

These results indicate that the mentoring group showed increases on the program quality scales (ITERS and Arnett). This increase is especially noticeable on the ITERS. Further, there was a decrease in program quality with the control group, going from a score of 137 to 132. On the Arnett scale the mentoring group increased greater than the control group (11 point increase versus a 7 point increase).

Although the above results did not reach statistical significance,

Table 2
ITERS

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	134	141	+7	ns
Control Group	137	132	–5	ns

Table 3
Arnett

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	29	40	+11	ns
Control Group	33	40	+7	ns

Table 4
KIDI

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	14	14	-0-	ns
Control Group	14	15	+1	ns

Table 5
Bloom

	Pre-Test	Post-Test	Change	Significance
Mentoring Group	73	74	+1	ns
Control Group	87	91	+4	ns

when specific subscales are analyzed several show significant differences (see tables 6 and 7). Several of the subscales on the ITERS and Arnett reached statistical significance with positive changes in routines (greeting/departing, meals/snacks, nap time, diapering/toileting, health/safety practice/policy) learning activities (eye-hand coordination, active physical play, blocks, pretend play, cultural awareness), sensitivity, and appropriate discipline for the mentoring group. The only statistically significant finding with the control group was in a negative change in interactions in which the scores decreased from pre-test to post-test. Paired t-tests were used in all of these analyses for Tables 6 and 7.

Table 6
Mentoring Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	36	41	.005
Listening activities	8	9	ns
Learning activities	28	31	.05
Interactions	13	13	ns
Adult needs	17	19	ns
Arnett subscales			
Sensitivity	26	31	.001
Appropriate discipline	7	9	.05

Table 7
Control Group

	Pre-Test	Post-Test	Significance
ITERS subscales			
Routines	41	42	ns
Listening activities	9	8	ns
Learning activities	29	31	ns
Interactions	15	13	.02
Adult needs	17	17	ns
Arnett subscales			
Sensitivity	28	31	ns
Appropriate discipline	6	7	ns

Discussion

These data demonstrate that the sites that were mentored improved on the ITERS and the Arnett. This is an encouraging result in that the intervention was only 4 months long. It is an important finding because the majority of mentoring projects in the past have utilized anecdotal evidence to demonstrate their effectiveness. Very few programs have conducted randomized trials of their interventions.

It is clear from the data that training/technical assistance interventions are needed in infant toddler programs because of the low scores on various program quality measures. It is also discouraging in that the control programs did not improve in which the ITERS went from 137 (pre-test) to 132 (post-test). This is a finding that will be monitored over time to see if this trend continues. Hopefully this was just an aberration in the data; however there does seem to be support when these data are compared to other studies (Iutcovich, Fiene, Johnson, Koppel, & Langan, 1997).

The public policy implications are that an intensive mentoring intervention of only four months can produce positive, although not statistically significant, changes in the overall quality of child care programs both globally and with caregiver interactions. Previous research (Johnson, 1994) has indicated that increasing the number of hours of training produces more developmentally appropriate behaviors in child care staff. Mentoring fits this model because it is an intensive one on one intervention in which the mentor and protégé are engaged in problem

solving activities to improve the overall quality of the interactions and environment of the child care program.

References

Behar, L. & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology*, 10, 601–610.

Breunig, R. & Bellm, D. (1996). *Early childhood mentoring programs: A survey of community initiatives*. Washington, D.C.: Center for the Child Care Workforce.

Clarke-Stewart, A. (1987). In search of consistencies in child care research. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 105–119). Washington, DC: National Association for the Education of Young Children.

Feagans, L. & Farran, D. (1979). *Adaptive Language Inventory*. Unpublished instrument. University of North Carolina, Chapel Hill.

Fenichel, E. (1992). *Learning through supervision and mentorship.* Washington, D.C.: Zero to Three: National Center for Infants, Toddlers and Families.

Fiene, R. (1995). Utilizing a statewide training system to improve childcare quality. *Child Welfare, 74*(6): 1189–1201.

Fiene, R. (1996). Using a statistical-indicator methodology for accreditation. In S. Bredekamp (Ed.), *A Decade of Accreditation*. Washington, DC: National Association for the Education of Young Children.

Galinsky, E., Howes, C., Kontos, S., & Shinn, A. (1994). *The study of children in family childcare and relative care: Highlights of findings*. New York, NY: Families and Work Institute.

Goelman, H. & Pence, A. (1987). Effects of childcare, family, and individual characteristics on children's language development: The Victoria day care research project. In D. Phillips (Ed.), *Quality in childcare: What does research tell us*? (pp. 89–104). Washington, DC: National Association for the Education of Young Children.

Griffin, A. & Fiene, R. (1995). *Childcare planning and evaluation*. Washington, DC: Zero to Three, National Center for Infants and Toddlers.

Harms, T. & Clifford, R. (1980). *The early childhood environment rating scale*. New York: Columbia University Teacher's College Press.

Helburn, S. (1995). *Cost, quality and child outcomes in child care centers*. Denver, CO: Center for Research in Economics and Social Policy, Department of Economics, University of Colorado.

Howes, C. (1987). Quality indicators in infant and toddler childcare: The Los Angeles study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us*? (pp. 81–88). Washington, DC: National Association for the Education of Young Children.

Hresko, W., Reid, D. K., & Hammill, P. D. (1981). *The Test of Early Language Development*. Austin, TX: Pro-Ed.

Iutcovich, J., Fiene, R., Johnson, J., Koppel, R., & Langan, F. (1997). *Investing in our children's future*. Erie, PA: Keystone University Research Corporation.

Johnson, J. (1994). *Child care training and developmentally appropriate beliefs and practices of child care employees in Pennsylvania*. Harrisburg, PA: Center for Schools and Communities.

Jorde-Bloom, P. (1988). Assess the climate of your center: Use the early childhood work environment survey. *Day Care and Early Education*, Summer 1988: 9–11.

Katz, L. (1994). Perspectives on the quality of early childhood programs. *Phi Delta Kappan*.

Kontos, S. & Fiene, R. (1987). Childcare quality, compliance with regulations, and children's development: The Pennsylvania study. In D. Phillips (Ed.), *Quality in childcare: What does research tell us?* (pp. 57–79). Washington, DC: National Association for the Education of Young Children.

Love, J., Schochet, S., & Meckstroth, R. (1986). *Are they in any real danger: What*

research does—and doesn't—tell us about childcare quality and children's well being. Princeton, NJ: Mathematica Policy Research, Inc.

Maternal and Child Health Bureau (1997). *Stepping stones to Caring for Our Children*, Washington, D.C.: Maternal and Child Health Bureau, Department of Health and Human Services.

Morgan, G., Costley, C., Genser, A., Goodman, M., Lombardi, J., & McGimsey, J. (1993). *Making a career of it: The states of the states report on career development in early care and education*. Boston: The Center for Career Development in Early Care and Education, Wheelock College.

NAEYC (1987). *Accreditation standards*, Washington, D.C.: National Association for the Education of Young Children.

Scarr, S., Eisenberg, E., & Deater-Deckard, R. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9:* 131–151.

ARTICLE

Child Care Health Consultation Improves Infant and Toddler Care

Rosemary Johnston, RN, BSN, MSN, Beth A. DelConte, MD, FAAP, Libby Ungvary, MEd, Richard Fiene, PhD, & Susan S. Aronson, MD, FAAP

ABSTRACT

Introduction: Many families enroll their infants and toddlers in early education and child care programs. The Pennsylvania Chapter of the American Academy of Pediatrics recruited 32 child care centers that care for infants and toddlers to be linked with a child care health consultant (CCHC).

Method: Project staff assigned the centers alternately to an immediate intervention or a 1-year delayed intervention (contrast) group. At entry into the project, and then 1 and 2 years later, an evaluator assessed center compliance with 13 standards for infants and toddler care selected from *Caring for Our Children: National Health and Safety Performance Standards* (3rd ed.). Project staff linked the Immediate Intervention centers with a CCHC in Year 1. In Year 2, in a crossover comparison, project staff linked Contrast centers with a CCHC.

Results: Working with a CCHC effectively improved compliance with some selected health and safety standards. J Pediatr Health Care. (2017) ■, ■-■.

KEY WORDS

Child care, child care health consultation, health and safety, infants and toddlers

INTRODUCTION

Nationally, about 48% of children younger than 3 years of age are enrolled in organized child care facilities (Laughlin, 2013). Early educators (child care staff) care for these children for many hours and many days. The quality of their care has lifelong impact on their physical, developmental, and social–emotional well-being (Garcia, Heckman, Leaf, & Padros, 2016).

In 2013, the Early Childhood Education Linkage System (ECELS), a program of the Pennsylvania (PA) Chapter of the American Academy of Pediatrics (AAP)

Rosemary Johnston, Infant Toddler Quality Improvement Project Coordinator, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Beth A. DelConte, Pediatric Advisor, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Libby Ungvary, Director, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA.

Richard Fiene, Research Psychologist, Research Institute for Key Indicators, Middletown, PA, and Affiliate Professor, Prevention Research Center, The Pennsylvania State University, University Park, PA.

Susan S. Aronson, Pediatric Advisor, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System, Media, PA, and Retired Clinical Professor of Pediatrics, The University of Pennsylvania and The Children's Hospital of Philadelphia, Philadelphia, PA.

This project was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under grant No. H25MC26235 Community-Based Integrated Service Systems. The information, content, and/or conclusions in this paper are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, or the U.S. Government.

Conflicts of interest: None to report.

Correspondence: Rosemary Johnston, RN, BSN, MSN, Infant Toddler Quality Improvement Project Coordinator, PA Chapter, American Academy of Pediatrics, Early Childhood Education Linkage System (ECELS), 1400 North Providence Rd., Ste. 3007, Rose Tree Corporate Center II, Media, PA 19063; e-mail: rljrmj@msn.com.

0891-5245/$36.00

Copyright © 2017 by the National Association of Pediatric Nurse Practitioners. Published by Elsevier Inc. All rights reserved.

http://dx.doi.org/10.1016/j.pedhc.2017.05.005

received a 3-year grant from the Maternal and Child Health Bureau (MCHB). The purpose of the grant was to "improve state infant/toddler [I/T] child care quality initiatives (Quality Rating and Improvement Systems [QRIS] and professional development)…" MCHB's grant required selection and promotion of 10 or more standards from a list provided by MCHB from *Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs*, 3rd ed. (*CFOC3*; AAP, American Public Health Association, & National Resource Center for Health and Safety in Child Care and Early Education, 2011).

Child care programs in PA's QRIS, called Keystone STARS, are ranked from the entry level at STAR 1 to STAR 4. To earn a rating, programs must comply with state regulations and meet the requirements listed for the designated STAR level on the PA Key Web site (www.pakeys.org). For a STAR 4 rating, a center that serves infants and toddlers must have scores at or above 5 (*good*) on the seven subscales of the Infant and Toddler Environment Rating Scale–Revised Edition (ITERS-R; Harms, Cryer, & Clifford, 2006). The Personal Care Routines subscale of the ITERS-R has some health and safety items. Scores in this subscale and on health and safety items in some of the other subscales are among the lowest scoring ITERS-R items in PA and elsewhere. This finding is reported by the PA Key Program Quality Assessment Team (2016) and by the authors of the ITERS-R (Harms and Cryer, personal communication, 2014).

Child care health consultants (CCHCs) use observation, education, collaborative decision making, coaching, and mentoring to achieve quality improvement in the QRIS (Zaslow, Tout, & Halle, 2012). CCHCs base their work on needs and feasible implementation. For more than a decade, published research has confirmed that child care health consultation is an effective approach to improving health and safety compliance with national child care standards (Alkon & Bernzweig, 2008; Alkon et al., 2008; Alkon, Bernzweig, Kim, Wolff, & Mackie, 2009; Alkon et al., 2014; Alkon et al., n.d.; Alkon, Sokal-Gutierrez, & Wolf, 2002; Banghart & Kraeder, 2012; Carabin et al., 1999; Crowley, 2006; Isbell et al., 2013; Moon & Oden, 2005; Organizational Research Services & Geo Education and Research, 2007; Pacific Research and Evaluation, 2007, 2008; Ramler, Nakatsukasa-Ono, Loe, & Harris, 2006; Roberts et al., 2000a, 2000b) Most of these studies did not specifically target care for infants and toddlers.

Published studies document the following specific improvements associated with involvement of a CCHC. Sanitation and hygiene reduced respiratory and gastrointestinal illness and days absent for illness among young children in group care (Carabin et al., 1999; Kotch et al., 2007; Roberts et al., 2000a, 2000b). Nationally recommended practices related to active play, nutrition, and food handling were adopted (Alkon et al., 2014). Policies and procedures accompanied by staff training reduced hazards and injuries (Kotch, 2002; Organizational Research Services & Geo Education and Research, 2007). Training about safe infant sleep positioning and the infant sleep environment reduced risk of sudden infant death syndrome (Moon & Oden, 2005). Better monitoring and tracking of immunization data in child care programs was associated with more children having up-to-date vaccine documentation (Alkon & Bernzweig, 2008).

The PA AAP established ECELS in 1989. ECELS maintains a CCHC Registry and regularly communicates with registered CCHCs to provide professional development, technical assistance, and tools to enable their implementation of the CCHC role. PA's CCHCs include private and public health service providers and health professionals who teach in academic settings. Funding for CCHC work is unpredictable, making recruitment, education, and retention of CCHCs challenging.

PA's child care regulations require that child care providers have documents showing that enrolled children are up to date with preventive health services recommended by the AAP, including "a review of the child's immunized status according to recommendations of the ACIP [Advisory Committee on Immunization Practices]" (PA Department of Human Services, 2008). This regulation is not enforced. Few providers use any reliable way to ensure that enrolled children are up to date. ECELS encourages child care centers to use a well-tested and routinely updated online software application called WellCareTracker™ (Weinburg, 2002) to check child health records for up-to-date routine preventive health services. It is described, demonstrated, and offered for subscription at www.wellcaretracker.org. Using WellCareTracker™ eases the burden for child care providers to comply with the regulation and remind families to obtain these services in a timely manner.

METHODS

Design

The PA AAP's MCHB-funded Infant-Toddler Quality Improvement Project (ITQIP) was conducted by ECELS using a randomly assigned clinical trial with a crossover comparison of centers assigned to an immediate intervention or delayed intervention (comparison) group. ECELS (a) assessed child care center practices related to I/T care for 13 selected *CFOC3* standards (AAP et al., 2011) and (b) assessed whether compliance with these practices improved when centers were linked with a CCHC.

Selection of the *CFOC3* standards addressed in ITQIP

With input from early care and education stakeholders, ECELS chose 13 *CFOC3* standards (AAP et al., 2011) from a list provided by MCHB (Box 1). The selection criteria were that the standard is (a) associated with the highest and most common risks of harm to I/T (AAP, American Public Health Association, & National Resource Center for Health and Safety in Child Care and Early Education, 2013), (b) measurable and amenable to improvement with technical assistance and professional development provided by a CCHC over a 12-month period, and (c) found by state inspectors to have a high level of noncompliance according to state data (PA Office of Child Development and Early Learning, 2010).

Evaluation plan

The evaluation plan is a classic randomly assigned crossover clinical trial. See Figure 1 for the evaluation plan logic model.

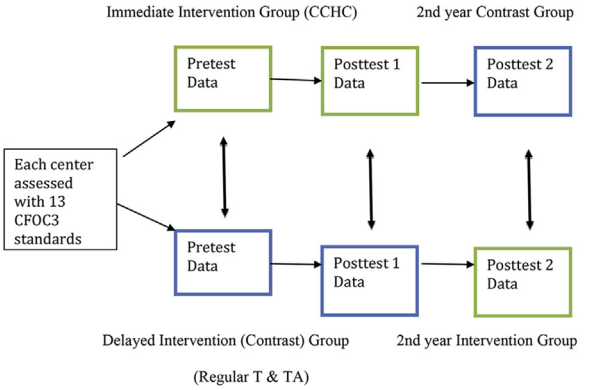
The ITQIP staff and consultants developed the evaluation tool described below. The ITQIP Project Coordinator (first author) and the evaluators collected data from participating centers at three points: when centers enrolled in the study (Pretest) and then 1 year (Posttest 1) and 2 years later (Posttest 2). One of the consultants (fourth author) compared the two groups on the pretest for equivalency and then on each of the two posttests. These data are discussed in the Results: Immediate Intervention Versus Delayed Intervention (Contrast) Group section. One year after the pretest data were collected, the participating centers were switched to a crossover comparison format. At this point, ITQIP ended the subsidy for the CCHCs who were working with the centers in the immediate intervention group and provided the subsidized CCHC linkage to the centers in the delayed intervention (contrast) group.

BOX 1. *CFOC3* standards chosen for ITQIP

1.4.5.2 Child Abuse and Neglect Education
3.4.4.1 Recognizing and Reporting Suspected Child Abuse, Neglect, and Exploitation
2.1.2.1 Personal Caregiver/Teacher Relationships for Infants and Toddlers
2.2.0.2 Limiting Infant/Toddler Time in Crib, High Chair, Car Seat, and other restraining equipment
3.1.3.1 Active Opportunities for Physical Activity
3.1.4.1 Safe Sleep Practices and SIDS Risk Reduction
3.2.1.4 Diaper Changing Procedure
3.2.2.1 Situations That Require Hand Hygiene
3.2.2.2 Handwashing Procedure
3.6.3.3 Training of Caregivers/Teachers to Administer Medication
3.5.0.1 Care Plan for Children with Special Health Care Needs
5.4.5.2 Cribs
7.2.0.1 Immunization Documentation

Note. CFOC3, *Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.); ITQIP, Infant-Toddler Quality Improvement Project; SIDS, sudden infant death syndrome*.

FIGURE 1. Evaluation plan logic model. CCHC, child health care consultant; T, training; TA, technical assistance.



When a center enrolled in ITQIP, the ITQIP coordinator interviewed the center director by phone. She gathered demographic data, including the number of enrolled I/Ts, where and when I/T activities occurred in the center, and the number of children who met the MCHB definition of special health needs. She asked the director to submit up to five of any care plans the center had for these children, redacted for confidentiality. The MCHB definition of a child with special health care needs is noted in *CFOC3* standard 3.5.0.1 as "a child who has or is at increased risk for chronic physical, developmental, behavioral or emotional conditions and who requires health and related services of a type or amount beyond that required by children generally" (AAP et al., 2011).

The ITQIP coordinator selected the rooms for the evaluator to observe as those with the largest number of children in the age group. The evaluators recorded observations in one infant and one toddler room at each center.

The evaluator collected a random sample of immunization records for up to 10 infants and 10 toddlers with the names redacted for confidentiality. The ITQIP coordinator used WellCareTracker™ software to check these immunization records. The ITQIP coordinator evaluated the care plans that the director submitted for the presence of the appropriate components from the list of the 14 components specified in CFOC3 standard 3.5.0.1. (AAP et al., 2011) and a 15th component, the presence of the health care provider's signature, that is required by PA regulations (Box 2).

The ITQIP coordinator scored the evaluator's observations of diapering, hand hygiene, and medication administration. She promptly prepared a summary of all the findings for the center and sent the summary to the center director and the linked CCHC before the first CCHC site visit. The summary delineated areas of strengths and areas to improve based on the evaluation tool results. To facilitate use of the data by the center staff and CCHCs, the summary included the text of the evaluation tool item, the center's score on the item, and the reason why the center met or did not meet the standard. The CCHC contacted the center within 2 weeks after receiving the summary to set up the initial site visit.

Evaluation Tool

The ITQIP staff prepared the items on the evaluation tool from performance guidelines specified in the 13 selected *CFOC3* standards (AAP et al., 2011). ITQIP consultants (fourth and fifth authors) and the ECELS staff reviewed the tool for clarity and validity of content. After several rounds of revisions, the ITQIP coordinator and a prospective ITQIP evaluator field-tested the tool, further revised it, and then field-tested it again, this time testing for interrater reliability with two evaluators independently and simultaneously using the tool.

The ITQIP evaluation tool has four sections: (a) Demographic Information collected in the phone interview (35 items), (b) Observations (64 items), (c) Interview Questions (28 items), and (d) Documents (14 items). The score awarded to items on the evaluation tool was based on the criteria listed in Box 3. A score of 2 or 3 for an item was considered a strength, and a score of 0 or 1 for an item was considered an area to improve. This total score was the sum of the scores for each item. The total number of scorable items on the evaluation tool is 106, with a maximum score of 318. The documents assessed include training records, written policies, care plans for children with special needs, immunization data, and PA child abuse clearances.

ITQIP assigned each scorable item to one of the 10 topic areas addressed by the 13 *CFOC3* standards selected for the project (AAP et al., 2011). See Table 1.

Sampling design: Recruitment, random assignment, and retention of centers

ECELS recruited Keystone STAR 2 and STAR 3 centers by distributing a flyer about the project. Programs with higher STARS ratings qualify for higher payments for children whose care is state subsidized. The highest payments are for children enrolled in STAR 4 centers. The increased payment for a higher rating is a quality improvement incentive. Also, ECELS offered participating centers three free $10 credit–awarding reviews for ECELS self-learning modules. The flyer was included in the newsletters of a variety of organizations: four of the five regional state-supported sources of professional development (Regional Keys), the PA Child Care Association, the Pittsburgh Association for the Education of Young Children, and United Way. Because the northwestern region of the state has the fewest centers, recruitment from that region was not attempted.

As the centers joined ITQIP, the project coordinator assigned them alternately to one of two groups, either the immediate intervention group or the delayed intervention (contrast) group. ITQIP enrolled centers from all four targeted regions of the state.

BOX 2. Care plan components evaluated for children with special needs

1. A list of the child's diagnoses
2. Contact information for the child's health care provider and any subspecialists
3. Medications to be administered on a scheduled basis
4. Medications to be administered in an emergency with clearly stated signs and symptoms in lay language about when to give the medication
5. Procedures to be performed while in care
6. Allergies
7. Diet modification that the child requires
8. Activity modifications
9. Environmental modifications
10. Triggers that cause a reaction to avoid
11. Symptoms for caregivers/teachers to observe
12. Behavioral modifications beyond those needed for a typically developing child
13. Emergency response plans for a facility emergency and if the child has an emergency event
14. Special skills training and education required and provided for the staff
15. Health care provider signature (required by Pennsylvania regulation)

Note. *Fourteen components specified in the* Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs *(3rd ed.) standard 3.5.0.1.* (American Academy of Pediatrics et al., 2011) *and a 15th required by Pennsylvania child care regulation*.

BOX 3. Criteria for scores assigned to items on the evaluation tool

0 = Never meets item
1 = Partly (<50%) meets item
2 = Mostly (≥50%) meets item
3 = Fully (100%) meets the item
NA = Not Applicable
NOp = Not observed or no opportunity to obtain data
DK = Don't know (interviewee response)

TABLE 1. Topic areas and number of items to score per topic

Abbreviation	Topic areas	Number of items to score per topic[a]
CA	Preventing Child Abuse	13
PR	Personal Relationships	9
LA	Limited Physical Activity of Infants	3
AO	Active Opportunity for Physical Activity	22
SS	Safe Sleep Practices/SIDS Prevention	19
MA	Medication Administration	8
DC	Diaper Changing Procedure	16
HH	Hand Hygiene	8
IM	Immunization Documentation	3
SN	Care Plans for Children With Special Needs	5

Note. *SIDS, sudden infant death syndrome.*
[a]*See the narrative for an item-by-item explanation of those items with significance levels (*p *values) based on the* t *tests performed on each item.*

Centers enrolled in ITQIP agreed to

- allow a 4- to 5-hour site evaluation once a year for 3 years,
- work with a CCHC for a period of 1 year to improve I/T health and safety,
- accept random assignment to one of the two project groups,
- provide access to redacted immunization records and care plans for evaluation,
- pay $240.00 of the $500 honorarium ITQIP paid to their CCHC, and
- remain in ITQIP for 3 years.

Recruitment and roles of evaluators and CCHCs

Evaluators.

ITQIP recruited 17 evaluators from the list of CCHCs who had previously received CCHC training from ECELS and from the nurses in the Maternal Infant and Early Childhood Home Visiting Program. All evaluators were health professionals with pediatric experience related to observed items. Most had experience working with *CFOC3* standards (AAP et al., 2011). The evaluators learned how to use the evaluation tool by participating in a live Webinar or by using the recording of the Webinar. All evaluators received a copy of the evaluation tool and a training manual with instructions for completing the evaluation. Seven evaluators were also CCHCs in this project. None of the evaluators who were CCHCs in ITQIP were linked with centers they evaluated.

The evaluators gave their completed evaluation tools to the ITQIP coordinator to score and summarize. The coordinator reviewed each submitted evaluation tool and then discussed the documentation with the evaluator by phone to make sure the scoring was as intended.

Child Care Health Consultants.

ECELS recruited 14 registered nurses and one physician as CCHCs. The ITQIP coordinator (first author) has worked as a CCHC for more than 15 years. She and the project's director and primary investigator, a pediatrician (second author) educated, coached, mentored, and supported the work of the CCHCs. The CCHCs participated in a Webinar about the project scope and the use of the selected *CFOC3* standards (AAP et al., 2011). They received a training manual that included the 13 selected *CFOC3* standards (AAP et al., 2011) and resources to support best practice in each of the 10 topic areas. ITQIP provided additional resources and periodic *CFOC3* updates (AAP et al., 2011).

During the site visit, the CCHC compared her observations with those in the summary and solicited concerns about health and safety practices from the center's staff. Then the director, program staff, and CCHC chose three of the 10 topics as the primary focus of the center's improvement. The CCHC helped the center staff prepare an action plan to work on the three topic areas they chose.

> The CCHC helped the center staff prepare an action plan to work on the three topic areas they chose.

Action plans included filling gaps in knowledge, developing policies for staff and family handbooks, and improving staff practices. The CCHCs and center directors arranged all subsequent contacts and visits over the next 12 months.

Quarterly, the CCHCs sent the ITQIP coordinator documentation of their work and progress toward goals. The CCHCs submitted the center's initial action plan and a final action plan at the end of the year that showed what the center accomplished. ITQIP paid $250 to the CCHCs upon receipt of the center's initial action plan and date of the first CCHC visit. ITQIP paid the CCHCs an additional $250 after they submitted the final action plan from their 12-month linkage. Throughout the project, the ITQIP coordinator reviewed quarterly encounter forms that the CCHCs submitted to describe their work with the centers. This enabled the ITQIP coordinator to suggest ways to promote progress on action plans, including use of relevant health and safety resources.

RESULTS

Descriptive Report

ITQIP linked CCHCs with 32 centers. Of these, 16 centers were in the immediate CCHC-linked group, and 16 were in the delayed CCHC-linked group. In all,

59 directors, 348 I/T teachers and 1,490 infants and toddlers were directly involved in ITQIP. Three centers from each group dropped out, leaving 13 centers in each group at the completion of the project (Table 2).

Over the 1-year period of CCHC linkage, 12 of the 32 programs had turnovers of two to four directors. This change in center leadership made the CCHCs' work to improve I/T care very difficult. For the immediate intervention group, three of the original 16 centers withdrew from the project. One center in the delayed intervention (contrast) group closed during the project period; two others withdrew from ITQIP. Some centers dropped out because they were so overwhelmed with maintaining ratios in classrooms and staffing issues that they believed they could not focus on their action plans.

This report compares pretest, Posttest 1 and Posttest 2 scores for the 13 immediate intervention sites and 13 delayed intervention (contrast) sites that remained enrolled in ITQIP for the full 3 years.

ITQIP did not require a specific time spent in the CCHC role for each linkage. The CCHCs in the immediate intervention group provided an average of 14 hours of consultation per site (range = 2.25–28.75 hours). The CCHCs in the delayed intervention (contrast) group provided an average of 12.5 hours of consultation per site (range = 2–32 hours). The CCHCs completed quarterly encounter forms to report the total hours of services to their linked center, including a checklist of onsite, phone, and e-mail services. The most common CCHC interactions with centers included providing health education for the director and staff, onsite consultation at the facility, technical assistance by phone or e-mail, providing print or audiovisual materials, helping the facility comply with state regulations, and developing health policies and procedures.

Topics chosen by the centers in the immediate intervention group and the delayed intervention (contrast) group and the number of centers that chose each topic are shown in Table 3.

Quantitative Comparison of Evaluation Tool Scores on the Pretest Versus the Two Posttests

The scores used in the quantitative comparisons are the sum of all scores on the Evaluation Tool, not only those for the topics that the center chose for special focus (Table 4).

Immediate intervention group

On the pretest, the range in scores was 175 to 267, with an average score of 212 out of a possible 318 points (66%). On Posttest 1, the range in scores was 213 to 297, with an average score of 254 out of a possible 318 points (79%). This change from the pretest to Posttest 1 was statistically significant ($t = -4.62$, $p < .0001$). Postest2 did not show any significant change from the average score on Posttest 1, showing that the initial results from the intervention were sustained in the next year (254 to 254).

Delayed intervention (contrast) group

On the pretest, the range in scores was 164 to 271, with an average score of 218 out of a possible 318 points (68%). On Posttest1, the range in scores was 149 to 257, with an average score of 221 out of a possible 318 points (69%). These changes from the pretest to Posttest 1 were not significant. Posttest2 showed significant change in the average score from Posttest 1 (221 points) to Posttest 2 (243 points; $t = -1.80$, $p < .08$) a year after this delayed intervention (contrast) group had received their CCHC linkage.

Immediate Intervention Versus Delayed Intervention (Contrast) Groups

The comparison of the average scores between the Immediate Intervention (212) and Delayed Intervention (Contrast, 218) groups on the pretest was not significant, showing that the groups were equivalent. The difference between the average scores of the immediate intervention (254) and delayed intervention (contrast, 221) groups on Posttest1 was statistically significant ($t = -3.46$, $p < .002$), showing the effectiveness of the CCHC intervention for the immediate intervention group. Posttest 2 showed no significant difference between the change in the average postintervention scores for the immediate intervention group 12 months after their CCHC-subsidized linkage and the delayed intervention (contrast) group (254 vs. 243) at the end of their 12 months of CCHC-subsidized linkage. See Figure 2 for the crossover comparison results.

TABLE 2. Location and retention of recruited centers

Region of Pennsylvania	Immediate intervention group			Delayed intervention group		
	Centers recruited	Centers dropped out	Centers completed	Centers recruited	Centers dropped out	Centers completed
Southwest Region (Pittsburgh metropolitan area)	1	0	1	3	1	2
South Central Region (Harrisburg metropolitan area)	4	1	3	2	1	1
Northeast Region (Allentown/Bethlehem/Scranton)	3	0	3	4	0	4
Southeast Region (Philadelphia metropolitan area)	8	2	6	7	1	6
Total	16	3	13	16	3	13

TABLE 3. *CFOC3* topics chosen by centers by intervention group

CFOC3 topics	Number of centers in immediate intervention group that chose each topic	Number of centers in delayed intervention (contrast) group that chose each topic
Safe Sleep Practice	11	11
Medication Administration	10	6
Child Abuse Prevention	6	1
Care Plans for Children with Special Needs	5	8
Diaper Changing Procedure	4	4
Limited Physical Activity of Infants	2	1
Hand Hygiene	2	5
Immunization	1	0
Personal Relationships	0	1
Active Opportunity for Physical Activity	0	4

Note. CFOC3, *Caring for Our Children: National Health and Safety Performance Standards; Guidelines for Early Care and Education Programs (3rd ed.).*

The crossover comparison results (Figure 2) show the relationship between the immediate intervention and the delayed intervention (contrast) groups in a crossover design. It clearly shows how effective the intervention (pretest to Posttest 1) was for the immediate intervention group and that the effects persisted after 1 year without a subsidized CCHC linkage (Posttest 1 to Posttest 2). It also shows that the intervention was effective when the delayed intervention (contrast) group was switched to receive the CCHC intervention with targeted training, technical assistance, and collaborative consultation a year after their pretest assessment (Posttest 1 to Posttest 2).

For the Immediate Intervention Group After 1 Year of Linkage With a CCHC

Among the items in each topic area (Table 1), the following items showed statistically significant improvement (pretest to Posttest 1).

Medication administration

The director had documentation that the staff who are authorized to give medications have received medication administration training within the year from a health professional ($p < .001$).

Safe sleep

The number of written safe sleep policies containing the required elements increased ($p < .05$). Teachers ($p < .01$) and parents ($p < .05$) reviewed the safe sleep policies and were educated about safe sleep practices ($p < .05$).

Child abuse

Child abuse policies contained the required elements ($p < .05$). Both infant and toddler teachers were educated about child abuse and how, as mandated reporters, they are required to personally report incidents they suspect might involve child maltreatment ($p < .001$). The number of centers having required clearance documents on file for teachers increased ($p < .05$).

Active opportunities for physical activity

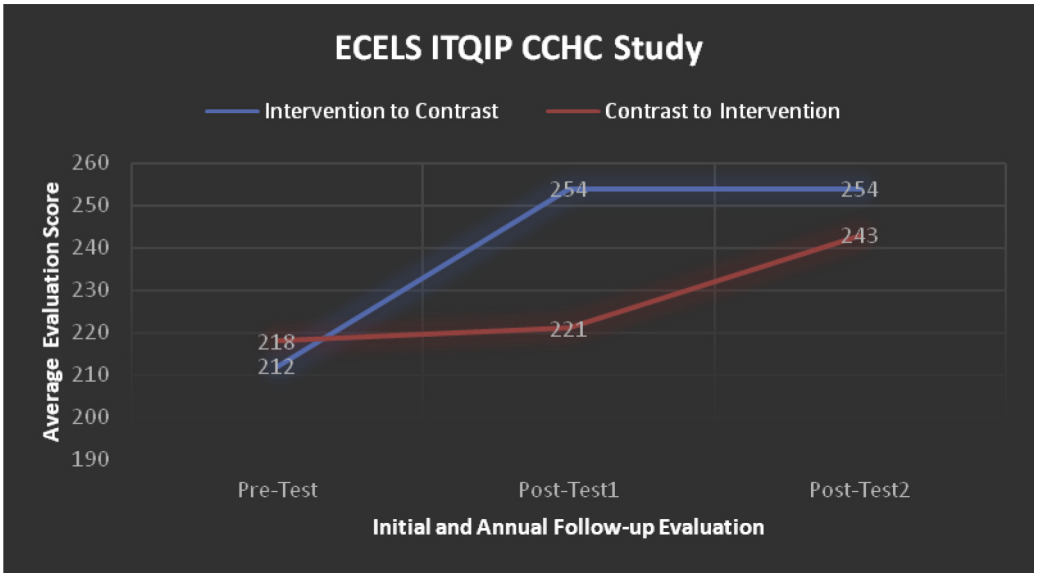
Infants (birth through 12 months of age) were taken outside two to three times per day, as tolerated ($p < .05$). Toddlers (12 months through 3 years)

TABLE 4. Quantitative results of the evaluation from the pretest to two posttests

	Intervention group				Delayed intervention (contrast) group			
	Range	Average	%	Possible total	Range	Average	%	Possible total
Pretest	175–267	212[a]	66	318	164–271	218	68	318
Posttest 1	213–297	254[a,c]	79	318	149–257	221[b,c]	69	318
Posttest 2	137–286	254	79	318	170–283	243[b]	76	318

Note. CCHC, child care health consultant.
[a]*Statistically significant change (*t = −4.62, p ≤ .0001*) from pretest to Posttest 1 for the immediate intervention group after the intervention of a 1-year linkage with a CCHC.*
[b]*Statistically significant change (*t = −1.80, p ≤ .08*) from Posttest 1 to Posttest 2 for the delayed intervention group after the intervention of 1 year of CCHC linkage.*
[c]*Statistically significant change (*t = −3.46; p < .002*) for Posttest 1 between the immediate intervention group and the delayed intervention (contrast) group.*

FIGURE 2. Crossover comparison results. CCHC, child care health consultant; ECELS, Early Childhood Education Linkage System; ITQIP, Infant-Toddler Quality Improvement Project.



went outside except in weather that poses a significant health risk ($p < .05$).

Diaper changing

Before the beginning of the diaper change, changing table paper was placed over the diapering surface, followed by the gathering of supplies needed for the diaper change from the containers in which they are stored and use of gloves ($p < .05$).

Hand hygiene

Observed times when toddlers ($p < .01$) and the toddler teachers/caregivers ($p < .05$) should have washed their hands showed statistically significant improvement after CCHC linkage.

For the Delayed Intervention (Contrast) Group After 1 Year of Linkage With a CCHC

Among the items in each topic area (Table 1), the following items showed statistically significant improvement (Posttest 1 to Posttest 2).

Safe sleep

Safe sleep policies that contained all the elements that should be in a safe sleep policy per *CFOC3* standard 3.1.4.1. ($p < .05$; AAP et al., 2011). The facility had documentation that parents reviewed the center's safe sleep policy and were educated about safe sleep practices ($p < .05$). There was no soft or loose bedding or other objects in a crib when an infant was in the crib ($p < .05$). Caregivers and teachers checked on sleeping infants often enough (about every 5 minutes) to be sure that the infant was still breathing ($p < .05$).

Medication administration

The name of a child to receive medication was verified before the medication was administered to that child ($p < .05$).

Diaper changing

Bottom clothing was removed, including shoes and socks, if feet were unlikely to be kept from contacting soiled skin or surfaces. If clothing was soiled, it was removed and placed in a plastic bag ($p < .05$).

Special needs

The number of care plans submitted that included the required elements in a care plan for children with special needs per the *CFOC3* standard 3.5.0.1 increased ($p < .05$; AAP et al., 2011).

Additional Findings of Interest

Immunization documentation

Only one center chose to work on documentation of up-to-date immunization status as an action plan focus. Overall, the immunization data for the two groups showed low compliance with *CFOC3* standard 7.2.0.1 (AAP et al., 2011) and PA's immunization regulations (PA Department of Human Services, 2008). On the pretest, in the immediate intervention centers, 22% of the immunization records for infants and 43% of the immunization records for toddlers were up to date.

Little change occurred for this group on Posttest 1 (36% for infants, 43% for toddlers.) On the pretest for the delayed intervention (contrast) centers, 25% of the immunization records for infants and 40% of the immunizations records for toddlers were up to date. On Posttest 1 the delayed intervention (contrast) centers improved from 25% to 38% for infants but dropped from 40% to 27% of the records for toddlers showing up-to-date vaccines.

Care plans for children with special needs

The data for the two groups showed low compliance with *CFOC3* standard 3.5.0.1 (AAP et al., 2011) that lists the components for care plans. Combining the immediate intervention and delayed intervention (contrast) center findings for this topic, the pretest showed that 66 I/Ts were identified with special health care needs in the 32 centers initially enrolled in ITQIP. Only 15 (23%) of I/Ts with identified special health care needs had any care plan signed by a health care professional. Only 1 of 66 I/Ts with special health care needs had a care plan signed by a health care professional that had all necessary components for optimal daily and/or emergency care. Posttest 2 showed that 39 I/Ts were identified with a special health care need in the remaining 26 centers. For children identified by the centers as having a special health care need, 62% did not have a care plan. Fifteen (38%) of those with identified special health care needs had a care plan signed by a health professional. Four of the 15 care plans had all the required elements. Examples of children who had special needs and had no care plan signed by a health care provider included children with gastroesophageal reflux taking Ranitidine, febrile seizures, asthma, multiple epinephrine autoinjectors onsite, autism, nonfebrile seizures, and torticollis and plagiocephaly, which required that the child wear a helmet each day.

DISCUSSION AND CONCLUSIONS

Quality early education and child care have been shown to be associated with lifelong benefits (Garcia et al., 2016). Young children are especially vulnerable to infectious diseases and injuries because of their age-appropriate behavior and abilities, their immature immune systems, and their lack of understanding of risk. Maintaining safe and healthful environments and practices involves removal of hazards and provision of policies and procedures, as well as compliance with quality standards by everyone in the group.

Numerous studies have shown the effectiveness of child care health consultation. This study focused on I/T care. The immediate intervention group showed significant improvement in policy development for safe sleep and child abuse and in education about safe sleep practices, preventing child abuse, and medication administration training. Some improvement in diaper changing and hand hygiene procedures occurred. The delayed intervention (contract) group showed significant improvement in safe sleep procedures, policies and education, medication administration procedure, diaper changing procedures, and care plans for children with special needs with appropriate information and signed by a health care provider.

The immediate intervention group showed significant improvement in policy development for safe sleep and child abuse and in education about safe sleep practices, preventing child abuse, and medication administration training.

The data collected by ITQIP show that many children with special needs lacked appropriate care plans. After finding little improvement in the immediate intervention group for centers having care plans with needed elements, ITQIP chose this topic as the focus of an MCHB-required continuous quality improvement initiative. ITQIP provided an audioconference for the CCHCs and gave them resources for teaching what should be in a care plan. CCHCs reported that they were most successful at helping the centers have complete, useful care plans for children with disease-specific conditions.

The areas chosen to target varied from center to center. Immunization was chosen by only one center. At the time of the study, neither regulation inspectors nor quality rating assessors were checking whether the center had documentation that the enrolled children were up to date with their vaccines. With little incentive or sanctions, documentation of up-to-date immunization status was poor.

Improvements occurred in some practices specified in selected *CFOC3* standards. Many of the directors said they appreciated the help they received from the CCHCs that ITQIP linked with their centers. The director of one center, part of a corporation with centers in 12 states, advocated for improving sleep policies for all the centers in her company. This advocacy could lead to widespread improvement.

Many of the directors said they appreciated the help they received from the CCHCs that ITQIP linked with their centers.

The centers that participated in this project were STAR 2 and STAR 3 programs that responded to an invitation to participate in ITQIP to improve. They were willing to contribute a modest copayment to work with a CCHC and wanted to raise their STAR rating and consequent higher payments for subsidized enrollees.

This selection bias is likely to have influenced the observed improvements.

A limitation of the study is the small sample size due to limited funding for the project. Also, although the study assessed practices for 13 *CFOC3* standards (AAP et al., 2011), the centers addressed only three topic areas. Little improvement was seen in topics that were not chosen or chosen less frequently. Change in leadership at the centers with varying levels of interest in working on the action plans made improvement difficult.

Another limitation of the study is the variability in child care operation from one facility to another and from year to year. Evaluators were unlikely to have been evaluating the same children from pretest through Posttest 2. Different teachers/caregivers and children may occupy designated rooms in a facility. ITQIP did not require that the CCHCs spend a specific amount of time with their centers. The time and type of service provided by CCHCs varied widely. Although CCHCs reported the total time and types of services they provided, they were not asked to report the time spent in each type of service (onsite visits, phone calls, or e-mails).

CCHCs support health and safety practices and environments that prevent harm and promote health and development of children, as well as overall wellbeing for families and early education staff. Currently, only 17 U.S. states have a statutory requirement for early childhood education programs to have child care health consultation (Honigfeld, Pascoe, Macary, & Crowley, 2017). Of these, two states require CCHC involvement only if the facility cares for sick children (Honigfeld et al., 2017).

None of the centers in this project continued their relationship with their CCHC after the year of subsidized linkage. Some directors stated that although they found the CCHC very helpful and informative, the cost of the CCHC was prohibitive. Some said they would continue the CCHC on a fee basis if they could budget for it in the future. Other studies have shown that linkage of centers with CCHC improves health and safety compliance. ITQIP showed this is true for I/T programs, too.

REFERENCES

Alkon, A., & Bernzweig, J. (2008). *Child care health linkages project evaluation summary. California Childcare Health Program, UCSF School of Nursing, September 2001 – June 2004*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1525-1446.2008.00689.x/full

Alkon, A., Bernzweig, J., Kim, T., Mackie, J., Wolff, M., & Elman, J. (2008). Child care health consultation programs in California: Models, services, and facilitators. *Public Health Nursing, 25*, 126-139.

Alkon, A., Bernzweig, J., Kim, T., Wolff, M., & Mackie, J. (2009). Child care health consultation improves health and safety policies and practices. *Academic Pediatrics, 9*, 366-370.

Alkon, A., Crowley, A., Benjamin Neelon, S., Hill, S., Pan, Y., Nguyen, V., …, Kotch, J. B. (2014). Nutrition and physical activity randomized control trial in child care centers improves knowledge, policies, and children's body mass index. *BMC Public Health, 14*, 215.

Alkon, A., Nouredini, S., Swartz, A., Sutherland, A. M., Stephens, M., Davidson, N., & Rose, R. (2016). Intergrated Pest Management Intervention in child care centers improves knowledge, pest control, and practices. *Journal of Pediatric Health Care*, *30*, e27-e41.

Alkon, A., Sokal-Gutierrez, K., & Wolff, M. (2002). Child care health consultation improves health knowledge and compliance. *Pediatric Nursing, 28*, 61-65.

American Academy of Pediatrics, American Public Health Association, National Resource Center for Health and Safety in Child Care and Early Education. (2011). *Caring for our children: National health and safety performance standards; Guidelines for early care and education programs* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics; Washington, DC: American Public Health Association.

American Academy of Pediatrics, American Public Health Association, National Resource Center for Health and Safety in Child Care and Early Education. 2013. *Stepping stones to caring for our children: National health and safety performance standards; Guidelines for early care and education programs* (3rd ed.). Elk Grove Village, IL: American Academy of Pediatrics; Washington DC: American Public Health Association.

Banghart, P., & Kraeder, J. L. (2012). *What can CCDF learn from the research on children's health and safety in child care? Urban Institute Brief #03*. Washington, DC: Urban Institute. Retrieved from http://www.urban.org/publications/412579.html

Carabin, H., Gyorkos, T., Soto, J., Joseph, L., Payment, P., & Collet, J. P. (1999). Effectiveness of a training program in reducing infections in toddlers attending day care centers. *Epidemiology, 10*, 219-227.

Crowley, A. (2006) *Improving child and family health through child care health consultation. Executive summary*. Beatrice Renfield-Yale School of Nursing Clinical Research Initiatives Fund. Yale University School of Nursing. New Haven, CT: Author.

Garcia, J. L., Heckman, J. J., Leaf, D. E., & Padros, M. J. (2016). *The life-cycle benefits of an influential early childhood program. Working paper 2016-035*. Chicago, IL: University of Chicago. Retrieved from https://heckmanequation.org/resource/research-summary-lifecycle-benefits-influential-early-childhood-program/

Harms, T., Cryer, D., & Clifford, R. (2006). Infant/toddler environmental rating scale (revised ed.). New York: Teachers College Press.

Honigfeld, L., Pascoe, T., Macary, S., & Crowley, A. (2017). Promoting children's health in early care and education settings by supporting health consultation. *Impact*. Farmington, CT: Child Health and Development Institute of Connecticut.

Isbell, P., Kotch, J., Savage, E., Gunn, E., Lu, L., & Weber, D. (2013). Improvement of child care programs' policies, practices, and children's access to health care linked to child care health consultation. *NHSA Dialog: A Research to Practice Journal, 16*, 34-52.

Kotch, J. B. (2002). *The Quality Enhancement Project for Infants and Toddlers: Executive summary*. Chapel Hill, NC: University of North Carolina at Chapel Hill.

Kotch, J. B., Isbell, P., Weber, D. J., Nguyen, V., Savage, E., Gunn, E., …, Allen, J. (2007). Hand-washing and diapering equipment reduces disease among children in out-of-home child care centers. *Pediatrics, 120*, e29-e36.

Laughlin, L. (2013). Who's minding the kids? Child care arrangements: Spring 2013. Current Population Reports, P70–135. Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/prod/2013pubs/p70-135.pdf

Moon, R., & Oden, R. (2005). Back to sleep: Can we influence child care providers. *Pediatrics, 112*, 878-882.

Organizational Research Services and Geo Education & Research. (2007). Child care health consultation: Evidence based effectiveness. Data from Healthy Child Care Washington evaluation report 2003–2007. Seattle, WA: Washington State Department of Health. Retrieved from https://www.napnap.org/sites/default/files/userfiles/membership/CCS_SIG_Evidence_%20Based_%20CCHP.pdf

Pacific Research and Evaluation. (2008). *Improving the health and safety of children in Oregon's child care: Implementation and outcomes of Oregon Child Care Health Consultation Program*. Portland, OR: Department of Human Services, Public Health Division, Office of Family Health.

Pacific Research and Evaluation. (2007). *Evaluation of the Child Care Health Consultation Demonstration Program: Phase IV Final Report* 2007. Portland, OR: Department of Human Services, Office of Family Health.

Pennsylvania Department of Human Services. (2008). 55 Pa Code, Chapter 3270, Child Care Regulations; 3270.131(a)(5). Retrieved from http://www.dhs.pa.gov/provider/earlylearning/childcareregulations/

Pennsylvania Key Program Quality Assessment Team. (2016). [2015–2016 Environmental rating scale reports: ECERS-R, ECERS-3, ITERS-R]. Unpublished raw data.

Pennsylvania Office of Child Development and Early Learning. (2010). *Demonstrating quality: Pennsylvania Keystone STARS 2010 Program report*. Harrisburg, PA: Author. Retrieved from http://www.ocdelresearch.org/Reports/Forms/AllItems.aspx?RootFolder=%2FReports%2FKeystone%20STARS&FolderCTID=0x01200092EA27E29EEE3E4AAE2D4C5508AC9E5A&View={5EEC6855-F8A8-486E-B6E0-FE6B9FDEBE2E}

Ramler, M., Nakatsukasa-Ono, W., Loe, C., & Harris, K. (2006). *The influence of child care health consultants in promoting children's health and well-being: A report on selected resources*. Rockville, MD: Maternal Child Health Bureau.

Roberts, L., Smith, W., Jorm, L., Patel, M., Douglas, R. M., & McGilchrist, C. (2000a). Effect of infection control measures on the frequency of upper respiratory infection in child care: A randomized, controlled study. *Pediatrics, 105*, 738-742.

Roberts, L., Smith, W., Jorm, L., Patel, M., Douglas, R. M., & McGilchrist, C. (2000b). Effect of infection control measures on the frequency of diarrheal episodes in child care: A randomized, controlled study. *Pediatrics, 105*, 743-746.

Weinburg, S. (2017). WellCareTracker™ [Computer software]. Retreived from www.wellcaretracker.org

Zaslow, M., Tout, K., & Halle, T. (2012). *On-site approaches to quality improvement in quality rating and improvement systems: Building on the research on coaching. Research-to-Policy, Research-to-Practice Brief OPRE 2012-40*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Pergamon

Journal of Early Childhood Teacher Education 22 (2001) 267–274

Journal of Early Childhood Teacher Education

Mastering course content and learner satisfaction in early childhood education: A comparison of regular classroom instruction with three variations of internet delivery☆

James Johnson[a,*], Richard Fiene[a], Jane Keat[a], Harriet Darling[a], Donald Pratt[b], Joyce Iutcovich[b]

[a]*The Pennsylvania State University, Pennsylvania, USA*
[b]*Keystone University Research Corporation, USA*

Abstract

In the spring semester 2000, a Penn State course, ECE 479 (The Young Child's Play as Educative Process), was taught by the same instructor in four delivery formats. One group consisted of a regular classroom, held on campus. A second group, also on campus, was taught in a computer lab via the Internet; and there were opportunities for interaction with peers and the instructor. A third group took the course on the Internet as part of a local distance education group; hence, there were some limited opportunities for face-to-face interaction with peers and the instructor. The fourth group took the course on the Internet, as part of a statewide distance education group, where there were no opportunities for face-to-face interaction. Twenty students who enrolled in the course (5 per group) completed questionnaires and phone interviews. Information was gathered on professional backgrounds, computer experience, and initial level of content knowledge on the topic of the ECE Internet course. Sixteen students who completed the course were interviewed again to evaluate satisfaction with the course and to estimate learning outcomes. Across the four conditions general satisfaction was expressed with the content, activities, and course requirements and with the teacher. However, students in the three computer groups expressed dissatisfaction over technical problems (all four who did not complete the course came from these computer groups). Significant gains in content knowledge occurred for the classroom group, while the learning in the three Internet-based instruction groups did not show the same gains. Concern was expressed related to the lack of face-to-face interaction, making the learning environment less desirable. Although Internet technology provides a great deal of promise, these results suggest that improvements are needed to make this delivery modality more effective for in-service distance learning. © 2001 Elsevier Science Inc. All rights reserved.

1. Introduction

"I think the thing that I liked least and I know it's inherent in an Internet course is lack of human involvement. Lack of human contact. Not getting to talk with the instructor at all, face to face. That was very tough for me. The other thing was the lack of the immediate feedback. If you had a question you had to type it in and wait for the instructor to log on and read it and then get back to you."

☆This research was funded by Keystone University Research Corporation and the Penn State Consortium for Children, Youth, and Families. The authors gratefully acknowledge the role played by ECE doctoral students Hey-Jun Ahn and Juli Graboski in helping develop the on-line course and World Campus specialist Cathy Holsing for her technical support and encouragement.

* Corresponding author. Tel.: +1-802-814-865 ext.2230; fax: +1-814-863-7602

E-mail address: jej4@psu.edu (J. Johnson).

0163-6388/01/$ – see front matter © 2001 Elsevier Science Inc. All rights reserved.
PII: S01090-1027(01)00129-5

"My frustration was that I don't like to fall behind in anything. I felt like I was falling behind because I wasn't able to access these readings, wasn't able to do some of the things on the computer that I should have been able to do. I was just getting frustrated and feeling like this isn't worth it for me. And that's why I dropped the course."

"The lack of being able to talk to anybody. The fact that you were doing all of this reading on your own. That was hard. I was used to sitting in classes and having someone lecture to me and then having students to interact with. It was difficult to get used to just reading it and absorbing it by yourself."

"Just simple things like just being able to print the page on my screen out. I didn't, I had no clue how to transfer stuff onto the hard drive and to be able to print something. I'm used to pressing the print button and stuff popping out. I didn't know it was that much involved."

The Internet is a new and ever expanding tool for learning (Benson & Meyers, 2000; Lan, 1999; Owston, 1998), including distance education (Williams, 1996). Whether or not the Internet can be an *effective* tool for training staff within the human services, particularly childcare staff, needs to be explored. Our project evaluated the effectiveness of the Internet training in terms of learning outcomes, its implementation (specifically, the technological aspects), and the student's level of satisfaction with the course. The impetus for our project came from two sources.

First, Pennsylvania delivers childcare training to all licensed and registered childcare providers in the state and is interested in making this system more cost effective and efficient. Discussions related to utilization of the latest technologies, such as the Internet, are being considered. Secondly, a new initiative, CyberStart, will link all licensed childcare centers in Pennsylvania to the Internet. While this initiative is specifically designed to offer Internet access and educational programming for children, it will also make this technology available to childcare staff. Hence, there is a need to evaluate the feasibility and effectiveness of Internet-based distance education as it becomes more available to childcare centers.

2. Research design

A quasi-experimental design was employed, which consisted of four groups of five students who enrolled in the early childhood education (ECE) course ECE 479 *The Young Child's Play as Educative Process,* which focused on play, communication, and curriculum. The first group experienced the traditional lecture/discussion course format. A second group took the course on the Internet, but within the context of a computer lab (located at Penn State York) where they had the opportunity to interact with their peers and the instructor. A third group took the course on the Internet as part of a local distance education (DE) group; this group also had some face-to-face interaction with their peers and knew the instructor. A fourth group took the course on the Internet, but as part of a statewide DE group. This group had no face-to-face interaction with their peers or the instructor.

This research design enabled us to examine the available technology to determine any hardware or software constraints, as well as the efficiency of the technological support services, by comparing groups that took the course via the Internet in different environmental circumstances (i.e., the on-campus computer lab vs. a home-computer set up). The research design also permitted us to evaluate the importance of the human element as a component of the effectiveness of this training modality since participants were in controlled settings with varying possibilities for face-to-face interaction.

Evaluating whether Internet training delivered to childcare staff is an effective training modality has special significance at this time given the paucity of empirical evidence to support or refute such claims. Research is needed to help the Pennsylvania Childcare and Early Childhood Development (PA CC/ECD) Training System, and comparable systems in other states, make more informed decisions about the use of Internet-based distance education for teachers of young children.

Given the exploratory nature of this study, a qualitative approach with a small sample was employed to generate data from questionnaires and interviews, as well as from course assignments completed by students. The questionnaire included items that tapped demographic characteristics (age, sex, and prior education), current position and experience within childcare, and experience with computers. Phone interviews, administered before and after the course, and lasting in duration from 30 to 45 min each, assessed students' knowledge about play and perceptions about the course. In addition, select course assignments were independently graded to assess knowledge.

The course aimed to increase the students' knowledge about play and its practical application. Two measures of the learning outcomes were used: (1) interview responses to questions about play given before and after the course and (2) grades on selected course assignments. The latter were independently evaluated by faculty other than the course instructor. Both measures were scored without knowledge of group membership.

From the phone interviews, students' answers to four questions about play were evaluated. The questions were: (1) What is play? (2) What is the value of play? (3) What is positive play? and (4) How can adults have a positive influence on play? Content analysis revealed that numerous ideas were elicited by these four questions both before and after the course. An empirically based coding system was developed and employed to score students' answers. Appendix A provides the categories and subcategories used in scoring.

3. Course development

The course *The Young Child's Play As Educative Process* (ECE 479), offered by the Department of Curriculum and Instruction in the College of Education at The Pennsylvania State University, was selected for this experimental study of Internet-based instruction. This particular course was chosen because it was viewed as well-developed and because it included both theoretical as well as practical content relating to ECE curriculum and instruction. Developing the course for Internet-based instruction required connecting with the Penn State World Campus (online learning). A contractual relationship was formed involving Keystone University Research Corporation (KURC), the contractor for the PA CC/ECD Training System, the College of Education, and the World Campus. Technicians from the World Campus, in collaboration with ECE faculty and doctoral students, worked to develop the course in time for offering it in the Spring of 2000.

3.1. Stage one – course structure

One of the first tasks was to establish the course syllabus, which included the course objectives and requirements, along with a schedule and sequence of learning activities. Objectives included acquiring a knowledge base of theory, research and practice concerning play and ECE, improving play observation and documentation skills, analyzing playthings for use in ECE settings, planning suitable indoor and outdoor environments for play, understanding and making sound judgments about the use of play facilitation strategies, and learning how and why to become an advocate for play in education. Requirements included an observation project, designing play environments, writing letters explaining play-based teaching to parents and to a 'blue ribbon committee' of educational professionals, as well as doing an implementation project and keeping a journal. *The Instructor's Manual to Accompany Johnson/Christie/Yawkey Play and Early Childhood Development, Second Edition* (Johnson,1999) provided the guidelines for the overall course organization and sequencing, which followed the chapters of the text, with the content going from theory and research to policy and practice.

3.2. Stage two – course content

The course was organized into four modules with a number of online lessons or sessions in each module. There were a total of 26 sessions. Sessions were designed to last 50 min; students had a reading assignment for each session and a self-administered objective-item exam, which produced computer-generated feedback for self-evaluation. Also, open-ended discussion questions were assigned for chat room or bulletin board discussion. The objective items and the open-ended discussion questions came from the instructor's manual. Activities and problems were embedded in sessions and were sequenced and used in the same way across the four groups of students.

3.3. Stage three – course programming

The final formatting and programming of the course for Internet instruction occurred during the Fall of 1999. The course program included some special features to make the on-line learning experience more interesting. For example, an animated pop-in character (a cartoon owl) appeared on screen at various selected points throughout the sessions to ask questions as a real classmate might. The World Campus technical staff, with the assistance of the course instructor (third author), also prepared a home page for the course, which included the course schedule and contact information.

4. Participants

All 20 students were female, ranging in age from 23 to 60 with mean of 39 years. Sixteen students had a bachelor's degree and one a master's and two had a high school diploma or GED certificate. The most common college major was elementary education ($N = 7$). Only three students majored in ECE and had taken any previous courses on play. For the entire group, inservice training hours on this subject ranged from zero ($N = 5$) to 45 hours ($N = 1$), with a mean of 10.5 hours.

Fifteen students held the position of center director and three were assistant directors. One student had the title of childcare coordinator, and another was a personnel training coordinator and an assistant group supervisor. The length of time that the students worked at their current childcare center ranged from

a minimum of six months to a maximum of 23 years with the mean 9 years. The length of time that students had worked in the early childhood field ranged from 2 years to 25 years, with mean 11.9 years.

5. Findings and discussion

5.1. Prior computer experience

Students reported prior experience using personal computers ranged from no experience to 15 years. The mean value was 4.1 years. Seventeen students had a personal computer at home and were asked to indicate which specific activities they had used. Those uses included: word processing (15 responses); recreational software (12 responses); Internet (12 responses); and spreadsheets and/or databases (4 responses).

A questionnaire item asked who usually provided technical support for their home PC (e.g., installing new software or hardware, answering software questions, and fixing problems). No more than five students usually relied on themselves for technical support for their home PCs. Of the four students who chose "other," three students usually relied on a friend, and one usually relied on another teacher in her childcare center.

5.2. Knowledge about play

For the following four play questions, each student's answers given in the postcourse interview were compared with the precourse answers in order to evaluate whether there were response improvements, defined by an increase in positive responses and/or a decrease in negative responses (items 1 and 3), by an increase in the total number of distinct acceptable benefits of play given (item 2), or of distinct teacher's roles (item 4) cited.

5.2.1. Question 1: what is the definition of play?

Responses were coded into the categories and subcategories and summed into two groups. *Positive responses* tallied the number of Attributes, Distinctions, and Realizations (A, D, R), while a second group of *negative responses* summed the total of overly-inclusive, mistakes, slogans, and vague replies (X, M, S, V). Other codes were considered "neutral" answers and were not used in this evaluation.

All classroom and computer lab students improved, as did the three statewide DE students who stayed in the course. Of the three remaining local DE group students, one failed to exhibit a positive response at both times of measurement; they gave five negative responses at precourse time of measurement and three negative responses at postcourse time of measurement.

Table 1
Pre- and post-course mean scores for benefits of play across four conditions

Groups	Classroom	Computer lab	Local DE	Statewide DE
Pre-course	5.2	4.0	3.8	3.0
Post-course	8.4	7.2	3.3	5.6

Classroom students gave a total of 25 positive responses after the course compared to only 3 positive responses before the course. This averages five positive responses per student and exceeds the averages on this index for the other three groups (1.2, 1.6, and 2.3 for lab, local DE, and statewide DE, respectively). Classroom and lab students showed few negative responses before or after the course, while local DE students gave 8 negative responses and statewide DE students gave 13 negative responses before the course. However, in postcourse interviews the six DE students made only five negative replies.

5.2.2. Question 2: what are the benefits of play?

Total scores for number of distinct benefits were derived from the coding system employed in the study. All general and expressive answers given (G and E) were added with all functional answers to yield a total benefits (G&E + F) score. Benefits of play results are shown in Table 1.

As can be seen in Table 1, the classroom and lab students performed better than the local DE and statewide DE students on this item. They were able to identify significantly more functions of play, especially after the course was taken. Nine of ten students in the classroom/lab groups improved, but only 3 of 6 in the DE groups did. Note also slight decrease in scores by local DE group.

5.2.3. Question 3: what is good play?

Here *positive responses* were the number of Nominal and Process (N and P) ideas from each student, while a second group of *negative responses* summed the total of Failure (F) responses exhibited. Other codes were considered "neutral" and were not used here.

Only seven of 16 students improved. Evidently, the course did not impact very much how well students could answer this question. The best answer, the P answer, was given by only three students, all at postcourse time. Failure to distinguish good play from play (e.g., "all play is good") happened six

Table 2
Mean scores for total roles across groups before and after the course

Groups	Classroom	Computer lab	Local DE	Statewide DE
Pre-course	3.8	2.4	4.0	3.6
Post-course	7.0	4.6	3.3	5.6

times at precourse time and eight times by the end of the course. Most students talked about types, functions and characteristics of play and failed to differentiate good play from play in an acceptable manner as determined by the coding and scoring system used in this study. This discouraging result may be because the scoring criteria were too harsh, or perhaps because interviewees did not understand the question. Also, it is possible that the course as taught, as well as the textbook as read, did not highlight the distinction well enough. In previous courses, this question has proven to be a "power question," distinguishing the "gold from the dross" in student performances.

5.2.4. Question 4: how can adults make child's play better?

Responses were analyzed with scores derived from the coding employed in the study. All general or attitudinal answers (G) per student were counted. Another group of answers covered the sum of all adult role responses (R). Total roles (G&R) were examined in the analysis. Table 2 shows mean scores across groups.

There were 11 students who improved, with all but one of them coming from the classroom or computer lab research groups. Only one student in a DE improved. Five students in DE groups actually performed more poorly in answer to this question after the course was over. In sum, the course influenced students' performance in identifying adult roles in children's play differentially depending on which research group they were in—classroom and lab students outperformed DE students.

5.3. Assessment of course requirements

There were three assignments that were graded for all the students: implementation activity (A), a parent letter (B), and a blue ribbon letter (C). As shown in Table 3, a comparison of the grades on these assignments across all four groups did not reveal that any one group consistently scored higher than the others. However, the traditional classroom group did score the highest on assignment B and had the second highest set of scores on assignments A and C.

Table 3
Scores on course assignments across four groups

Groups	Assignment A	Assignment B	Assignment C	Total
Classroom	2.86	3.86	2.75	9.47
Lab	1.90	2.63	3.00	7.53
Local DE	2.69	3.37	2.50	8.56
Statewide DE	2.99	3.61	2.35	8.95

For assignments A and B the differences in scores for the four groups were statistically significant (Assignment A, $F=5.574$, $p < .012$ and Assignment B, $F=4.628$, $p < .023$). No significant differences were found for Assignment C. The computer lab group scored the lowest on assignments A and B while the other three groups of students all scored about the same. When all the assignments are totaled for an overall score, the traditional classroom performed significantly better than the other three groups ($F=5.221$, $p < .015$).

Students were also asked a series of questions about their expectations regarding the course, the benefits from taking the course, and their experiences in taking the course. The majority of the responses were very positive about the course in general, but the students who had taken the course via the Internet experienced a number of computer problems in either getting online or with the software. Everyone experienced problems and this delayed the start of the course for them for several weeks. There were problems with signing on, with the chat room, with passwords not being accepted, and so forth. However, once these problems were worked out, the course started and proceeded fairly well.

In sum, respondents revealed general satisfaction with the content of the course, the course activities, and course requirements. Moreover, there was a strong appreciation and high evaluation for the teacher. On the other hand, almost everyone (one or two exceptions) in the three computer groups expressed dismay over serious and continuing technical difficulties throughout the entire semester. However, even with a good course content, instructional design, teacher, and technical delivery, there were several people who clearly indicated reservations about Internet learning because it lacked face-to-face interaction. These people indicated that even if there were no technical difficulties, they would miss the human contact and would prefer courses or training taken in a classroom where there was greater opportunity for interpersonal interaction and contact.

6. Conclusion

This evaluation provides insights into offering ECE courses over the Internet. Clearly, it seems that the success of this technology is dependent upon the persistence and knowledge of the student for learning to occur. Four students who did not complete the course were from the local ($N = 2$) and statewide ($N = 2$) Internet-based distance education groups; none were from the traditional classroom or computer lab settings. The students were administrators of childcare programs, fairly familiar with computers and knowledgeable of the course content. However, even with these pluses, the students still had considerable difficulty in accessing and doing the course online. Possibly the dropout rate would have been greater if the students were at a beginning stage of their career (Cohen, 2000; but see Schrum, 1992). Students in the traditional classroom and in the computer lab groups, where there was more face-to-face interaction, scored the best on the interview play evaluations. Evidently, these two training modalities were more effective than were the local and statewide distance education training modalities where there was little or no face-to-face contact with other students or with the instructor.

On the three course requirements, the classroom group scored the highest on assignments summed together (9.47); but the computer lab group scored the lowest on the assignments (7.53), with the two distance education groups in the middle. This is inconsistent with the results from the analyses of the interview responses to the four questions about play. Here both the classroom group and the computer lab group gained the most. Although it is perhaps encouraging to see that the students in the two distance education groups scored higher on the course assignments compared to the computer lab group, their results were still lower than the students' scores from the traditional classroom group.

What have we learned from this study? For Internet instruction to be effective it seems that students must be technologically literate and knowledgeable about course content to some degree, and they must be persistent and highly motivated. Without these personal characteristics, the Internet course experience may not be a positive learning experience. The results of this study further suggest not utilizing Internet training across-the-board for childcare staff. It needs to be very targeted, beginning with directors of programs, who generally have the greatest experience and education and potential exposure to computer technology. More generic, across-the-board training for the beginning level practitioner does not make training sense at this point.

Internet technology provides a great deal of promise for reaching childcare staff with needed specialized in-service training in ECE. But first it is necessary that the technology gets the fine-tuning to ensure its effectiveness as a training modality for the majority of childcare providers. Also, what is best for on-line learning? Professional attention must be given to criteria for deciding the kinds of training content (e.g. learning facts like state regulations or understanding and applying concepts like play in ECE) in relation to its packaging (e.g., a full course like in this study, or modules, or single sessions analogous to workshops). Setting up realistic expectations for learners and designing the right means of instruction and assessment are both very important. Finally, we recommend that priority be given to designing orientation modules to prepare potential users for Internet-based distance education.

References

Benson, M., & Meyer, J. (2000). Student teachers' reflections and collaborations through electronic mail discussions. *Journal of Early Childhood teacher Education, 21*, 123–127.

Cohen, L. (2000, April). *Are we ready to abandon the classroom? The dark side of web instruction.* Paper presented at the annual meetings of the American Educational Research Association, New Orleans.

Johnson, J. (1999). *Instructor's Manual to accompany Johnson, Christie, and Yawkey Play and Early Childhood Development* (2nd edition). New York: Addison, Wesley, Longman

Lan, J. (1999, March). *The impact of Internet-Based Instruction on Teacher Education:The "Paradigm Shift."* Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Washington, DC.

Owston, R. (1998). *Making the link: Teacher professional development on the internet.* Portsmouth, NH: Heinemann.

Schrum, L. (1992). Professional development in an informational age: An on-line experience. *Educational Technology, 32* (12), 49–53.

William, A. (1996, June). *Integrating courses with the internet: Preparing the teacher as well as the learner.* Paper presented at the summer conference meetings of the Association of Small Computer Users in Education, North Myrtle Beach, SC.

Appendix Table 1

Question 1: What is play?

A1 Attribute of play given. For example, play is process-oriented, marked by positive affect, often non-literal, active, intrinsically motivated, etc.
A2 Tangential attribute of play given. For example, play is something adults like to see in children; child and adult do play of own free will, etc.
E Example of play is given. For instance, play is like playing with dolls or blocks.
F Function of play is noted as a way to define it. A way to learn things. A way to release feelings.
C Context of play is noted. Play can be solitary or done in a social group.
X Overly inclusive or general statement about play is made, such as "it is a creative process."
D Distinction is made, contrasting play with other similar behaviors such as imitation or exploration.
R Realization is indicated, such as play is multidimensional, complex, or hard to define.
T1 Type of specific play is given, such as constructive or dramatic play.
T2 Type of general play is noted, such as educational play, dark play, and recreational play.
M Mistake is made, a falsehood is uttered.
S Slogan is cited; such as play is the child's work.
V Vague pronouncement is made, unclear or hard to decipher.

Question 2: What is the value of play? What are its benefits?

General (educational) values include:
G1 Teaches skills and abilities or is a learning experience
G2 School preparation
G3 Allows for the practice of skills
G4 Generates further learning and development; allows children to go as far as they can with what they are learning
G5 Avenue for creativity
G6 Creates a well-rounded child
G7 Allows child to explore and discover on one's own
G8 Allows child to experiment

Expressive values include:
E1 Enjoyment, fun, makes happy, love of life
E2 Relaxation, release energy, tension
E3 Vent frustration
E4 Be self
E5 Express self

Functional values include:
F1 Cognitive: abstract thinking * imagination * learning content * creativity * learn on own terms-relevancy, meaningfulness * problem-solving * meta-cognition * memory * social cognition, empathy, perspective-taking * theory of mind * sense of self * sense of others * assimilation, integration, application of learning
F2 Affective: motivating, feel good about self * self empowerment, sense of control * reduce anxiety, therapeutic, cathartic * self confidence, sense of self-assuredness
F3 Social: learn to resolve conflicts * cooperation * group cooperation, team member * leadership skills * learn to share * learn to take turns * learn to help
F4 Physical: gross motor * fine motor * learn to challenge self physically * self-help skills
F5 Attentional: concentration * attention regulation * persistence
F6 Assimilation
F7 Language: communication skills * literacy * become good story-teller * vocabulary
F8 Academic: reading and writing * shapes for math * science
F9 Life skills, careers

Question 3: What is good play? Positive play?

N Nominal answer is given, like educational play is good play.
P Process of play is said to be important, "perking along", playing up to capacity.
C Characteristic of play in general is given without really answering the question.
T Type of play in general is given without really answering the question.
F1 Function of play in general is given without really answering the question.
F2 Failure to provide sensible reply, such as, "all play is positive."

(*continued on next page*)

Appendix Table 1 (*continued*)

Question 4: How can adults make play better for children?

There are many roles to perform in the ECE profession with respect to children's play. Some are general or attitudinal such as:

G1 Value play, allow it to happen, be aware of your biases and those of others
G2 Realize when attitude change about play occurs in self or others
G3 Show an interest in play as a matter of public or educational policy
G4 Lobby and advocate on behalf of children's right to play as an integral part of childhood

Other roles in the ECE profession with respect to children's play pertain to classroom or home or child development center situations where young children learn and are cared for. These include:

R1 Stage manager, set up over-all positive and attractive environments, provide materials, playthings, space, toys, storage areas, time for play, literacy props, allow children choice of areas and activities, organize settings, rotate toys
R2 Be a careful and systematic observer, evaluate play, change own behavior as a result of observations, document play process and products
R3 Provide preparatory experience, bring in guest visitors, field trips
R4 Play facilitator, scaffold, support, challenge children at play, do not be overbearing, ask questions, offer comments, suggestions, ideas for play, teach play
R5 Co-player, be a play leader, model play behavior
R6 Supervise play, make sure play is safe, monitor activities, mediate conflicts, enforce rules, help children negotiate, guide and discipline, be a referee

Other roles inherent in the ECE profession connected with play have to do not with the children but with other adults. These include:

R7 Help adults be comfortable with play of children in educational settings, train teachers, educate staff about play in ECE
R8 Inform and involve parents in quality play with young children in formal and informal educational settings

RESEARCH REPORT

Penn State Better Kid Care Click. Coach. Connect.: Supporting Infant-Toddler Early Learning Project

Abstract

This research report provides the statistical and data analyses, and the Methodology for the Penn State's Better Kid Care Click. Coach. Connect.: Supporting Infant-Toddler Early Learning Project

Christine Anthony, Rebecca Escott, and Richard Fiene

Penn State's Better Kid Care Click. Coach. Connect.: Supporting Infant-Toddler Early Learning Project

Christine Anthony, Rebecca Escott, and Richard Fiene

The Pennsylvania State University

October 2019

The purpose of this report is to describe the efforts of an intensive infant-toddler, in-house coaching intervention – a hybrid model using digital resources and virtual support for onsite coaches through the Better Kid Care Program funded by the William Penn Foundation; as well as providing descriptive and demographic analyses of the Infant Toddler Environmental Rating Scale, Version 3 (ITERS-3). In starting any intervention study, it is paramount that one establishes equivalency between the intervention and comparison groups and usually there are always some very interesting descriptive and demographic trends that appear in the data. In this case because the ITERS-3 is so new, it is equally interesting to report on some very basic descriptive/demographic statistics drawn from this study so that other researchers can compare their respective samples with this sample. As the reader will see in the Results section, all of these relationships have been addressed.

Before beginning there are several interesting footnotes that need to be made to better understand the results of this study. Although the programs were randomly assigned to either the intervention group or the comparison group, all the programs that participated in this study were high performing programs as measured by the Pennsylvania Quality Rating and Improvement System Keystone Stars. They were all at either a Star 3 or 4 level which is indicative of a high performing early care and education program. In selecting the Comparison Group, it is not a typical Control Group in that it did have access to the Better Kid Care Online Modules, just the Onsite Coaching was not available; and the Comparison Group participants were asked not to engage in any coaching and did not receive any of the coaching supports available to the Intervention Group.

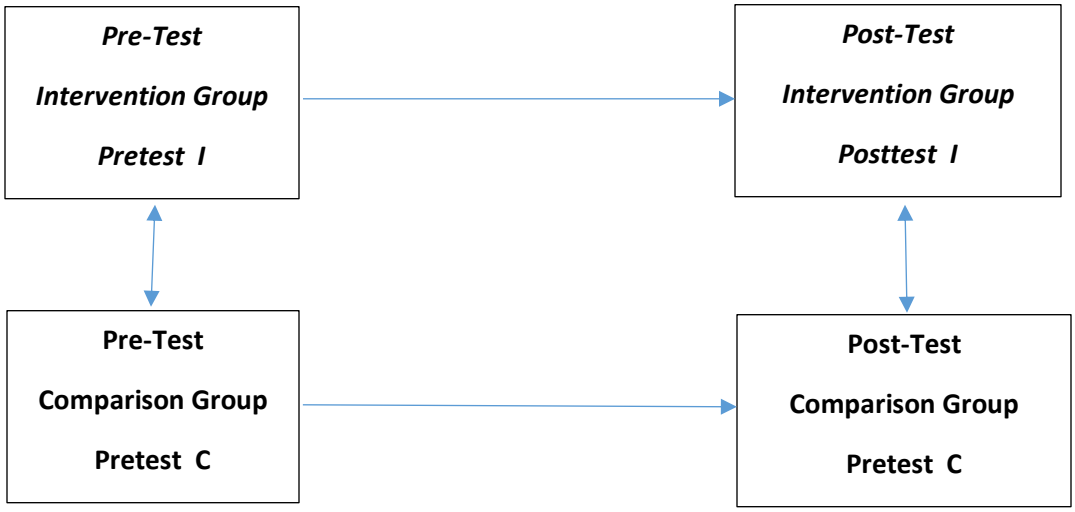
Coaching has been demonstrated to be a very effective professional development, training and technical assistance approach to improving the level of early care and education quality. The problem with coaching is that although it is very effective, it is costly to implement and therefore is not very cost efficient. This project is an attempt to combine the cost effectiveness of coaching with a cost efficient in-house delivery system supported by digital resources delivery system through the Penn State Better Kid Care project.

Coaching interventions are being used in many state professional development and training systems as well in state Quality Rating and Improvement Systems (QRIS) technical assistance in order to improve the overall quality of care being provided in early care and education programs. Interest in online approaches to training has grown over the past decade because of its cost efficiency and effectiveness and the ability to reach hard-to-reach early care and education providers, such as family child care. The Better Kid Care Click. Coach. Connect. Approach provided the following supports to the intervention group: a 300+ page hybrid learning guide aligned with the ITERS-3; full-day training; access to weekly coaching logs; and bi-weekly support calls with the project coordinator.

LOGIC MODEL FOR ANALYTICS

Figure 1 provides the logic model for the analytics for the study. It is a classic random clinical trial with an intervention group (Modules plus in-house coaching) with a comparison group (Modules only). Paired t-tests for the intervention group classrooms and comparison group classrooms were completed from pre- to post-tests. Independent t-tests were completed comparing the intervention group to the comparison group in both the pre- and post-tests. Correlational analysis was utilized to look at relationships amongst variables.

Figure 1 – Logic and Analysis Model:



METHODS

The focus of this study was Star 3 and Star 4 infant-toddler classrooms within the City of Philadelphia in Pennsylvania focusing on infant and toddler classrooms. There were 47 classrooms with 24 intervention classrooms and 23 comparison classrooms which began the project. Three observers collected the ITERS-3 data on the 47 classrooms. Basic demographic information was collected on each of the classrooms, their programs, teachers and directors, such as: profit/non-profit status, QRIS Keystone Star level (Levels 3 and 4 were included), years of ECE experience, years at present location, educational level of director, etc. Also, weekly coaching logs were completed by the coaches in the Intervention Group. There were 8 coaches in total. Data within the coaching log kept track of the observations made, the length of the coaching session, what was covered, and where it was conducted.

By the end of the intervention nine months later, there were 35 programs with 13 intervention classrooms and 22 comparison classrooms. It is unfortunate with the loss of the intervention classrooms (this will be discussed in the Limitations Section), but not unusual for this type of research.

RESULTS

The most salient result was the analyses between the Intervention and Comparison groups. There were no statistically significant differences between the two groups on the ITERS-3. The overall ITERS-3 score on the pre-test was 3.63 for the intervention group and 3.33 for the comparison group; and 4.57 and 4.19 respectively for the intervention and comparison groups on the post-test. But there were statistically significant differences from pre- to post-tests for both the Intervention and Comparison groups and the Effect Size was large for the Intervention Group (see Chart 1).

Chart 1 – Pre-Test & Post-Test ITERS-3 Scores for the *Intervention* and Comparison Groups

Groups	Pre-Test	Post-Test	Difference	Effect Size
Intervention (I) Group	***3.63***	***4.57***	***P < .01***	***1.21***
Comparison C Group	**3.33**	**4.19**	**P < .01**	**0.96**
Difference	**NS**	**NS**		
Effect Size	**0.36**	**0.46**		

Also, there were no statistically significant differences amongst the three assessors collecting the ITERS-3 data on either the pre- or post-tests (1 = 3.67, 4.62; 2 = 3.32, 4.02; 3 = 3.45, 4.54). All sub-scales and items within the ITERS-3 were non-significant on the pre-test and post-test. However, the t-tests did show significant differences for both the intervention and comparison groups in going from pre- to post-tests (see Charts 2a & 2b for the details related to the ITERS-3 Sub-Scale scores).

Chart 2a – Pre-Test Scores on the ITERS-3 for the *Intervention* & Comparison Groups

ITERS-3 Sub-Scales	*Intervention Group*	Comparison Group
Space and Furnishings	***3.87***	**3.17**
Personal Care Routines	***2.73***	**2.86**
Language and Books	***4.08***	**3.89**
Activities	***3.23***	**2.79**
Interaction	***4.28***	**4.05**
Program Structure	***3.37***	**3.24**

Chart 2b – Post Test Scores on the ITERS-3 for the *Intervention* & Comparison Groups

ITERS-3 Sub-Scales	*Intervention Group*	Comparison Group
Space and Furnishings	***4.75***	**3.93**
Personal Care Routines	***4.31***	**3.58**
Language and Books	***5.01****	**4.58***
Activities	***3.91****	**3.75**
Interaction	***5.13****	**4.91**
Program Structure	***4.70***	**4.36**

On the pre- to post-test comparisons it is evident that the intervention and comparison groups increased in both cases *(* = p < .05; ** = p < .01*). In comparing the Intervention Group to the Comparison Group, only the **Space and Furnishings** Sub-Scale demonstrated a statistically significant difference (I = 4.75; C = 3.93; p < .05), all other sub-scales were not statistically significant however the means were always higher with the Intervention Group.

Chart 3 – Item Scores for the *Intervention (I)* and Comparison (C) Groups Pre-Post-Tests[1]

Item[1]	*Pretest (I)*	Pretest C	*Posttest (I)*	Posttest C	*Difference I*	Difference C
Space 1	*3.54*	2.87	*3.23*	3.05	*NS*	NS
Space 2	*4.29*	3.74	*6.00*	4.23	**	NS
Space 3	*4.00*	4.00	*5.15*	4.86	*NS*	NS
Space 4	*2.79*	1.96	*4.46*	3.59	*	**
Personal 5	*2.33*	2.43	*4.46*	3.27	**	NS
Personal 6	*3.25*	3.39	*4.46*	3.86	*	NS
Personal 7	*1.75*	2.30	*3.85*	3.14	**	NS
Personal 8	*3.38*	3.17	*4.46*	4.05	*	NS
Language 9	*4.42*	4.30	*5.85*	5.23	*	NS
Language 10	*3.38*	3.74	*4.23*	3.77	*NS*	NS
Language 11	*3.96*	3.70	*5.38*	5.09	*	*
Language 12	*4.50*	4.57	*5.69*	5.09	*	NS
Language 13	*3.46*	3.39	*4.46*	4.23	*NS*	NS
Language 14	*3.42*	3.09	*4.46*	4.09	*	*
Activities 15	*3.25*	3.17	*4.62*	4.45	*	**
Activities 16	*2.00*	2.20	*3.08*	3.00	*NS*	NS
Activities 17	*3.00*	3.09	*3.54*	3.68	*NS*	NS
Activities 18	*3.33*	2.26	*3.54*	3.09	*NS*	NS
Activities 19	*3.08*	2.39	*4.39*	3.59	**	**
Activities 20	*3.25*	2.83	*4.92*	3.68	**	NS
Activities 21	*3.29*	2.65	*2.92*	3.36	*NS*	NS
Activities 23	*4.75*	3.83	*4.92*	5.36	*NS*	*
Activities 24	*2.08*	2.39	*3.54*	3.23	**	NS
Interaction 25	*3.92*	4.13	*4.92*	5.41	*NS*	**
Interaction 26	*4.38*	4.52	*5.54*	5.41	*	*
Interaction 27	*3.92*	3.78	*4.77*	4.50	*NS*	NS
Interaction 28	*4.00*	4.13	*6.00*	5.36	**	*
Interaction 29	*4.38*	4.35	*5.62*	5.09	*NS*	NS
Interaction 30	*3.38*	3.00	*3.92*	3.95	*NS*	*
Program 31	*2.92*	2.78	*4.85*	4.64	**	**
Program 32	*3.54*	3.65	*4.77*	4.18	**	NS
Program 33	*3.50*	3.20	*4.45*	4.21	*NS*	NS

[1] *See the Legend at the end of the report for a listing of each item.* (* p < .05; ** p < .01)

In comparing the Intervention Group to the Comparison Group with Post-Test Item Scores, two items were statistically significant, **Space 2 – Furnishings for Care Play and Learning** (I = 6.00; C = 4.23; $p < .01$) and **Activities 20 – Nature/Science** (I = 4.92; C = 3.68; $p < .05$).

In exploring the demographic, teacher, and coaching data elements and the pre- and post-test ITERS-3 data there were some significant differences between the ITERS-3 and these data. Chart 4 provides the basic demographic data displays for the total sample of both the intervention and comparison groups (see Chart 4). The other analyses amongst the demographics and ITERS are reported in the paragraph following Chart 4.

Chart 4 – Basic Demographic Information on Teachers, Programs, Directors, and Coaching

Number of Profit based Programs	**20**
Number of Non-Profit Programs	**15**
The Number of Years in the Specific Location for the Program (Average)	**23**
QRIS Keystone Stars 3 Program	**20**
QRIS Keystone Stars 4 Program	**15**
Licensed Capacity of the Program (Average)	**141 children**
Director's Years at Program (Average)	**9 years**
Directors with an AA Degree	**11%**
Directors with a BA Degree	**43%**
Directors with a MA Degree	**46%**
Lead Teacher's Years of Experience (Average)	**3 years**
Lead Teachers with a High School Diploma	**61%**
Lead Teachers with an AA Degree	**22%**
Lead Teachers with a BA Degree	**5%**
Pre-Test Average ITERS-3 Comparison Group	**3.33**
Pre-Test Average ITERS-3 Intervention Group	**3.63**
Post-Test Average ITERS-3 Comparison Group	**4.19**
Post-Test Average ITERS-3 Intervention Group	**4.57**
Number of Staff who completed at least 5 required modules	**60**
Percent of Time A Coaching Observation Occurred	**76%**
Percent of Time There Was a Focus for the Coaching Observation	**78%**

There was a significant relationship between ITERS-3 and the Keystone Stars level ($r = .31$; $p < .04$). There were significant relationships between profit vs non-profit status with the following: years in the location ($-.63$; $p < .0001$) and star level ($r = -.33$; $p < .03$) favoring non-profit status. There were statistically significant differences between star levels 3 and 4 (3.20 vs 3.76 respectively)($F = 4.71$; $p < .04$); and a trend for non-profit programs to score higher on the ITERS-3 (3.59) versus profit programs (3.17). There were no ITERS-3 score differences by education of director or teachers. This relationship has been found in previous studies.

There was another confounding staffing issue with turnover in both the Comparison and Intervention Groups. It is widely known that turnover rate in child care staff is a constant problem. The sample for this study was no different than what has been reported in the research literature with a 40+% in the Comparison Group and a much larger 60+% in the Intervention Group. This tremendous turnover (the number of lead teachers who changed between pre- and post-test), it was interesting to look at the ITERS-3 Scores for the two groups. The major finding was that the programs where there was no change in the teachers, the ITERS-3 scores were greater than those programs in which there was a change in teachers (4.41 vs 4.26 on the post-test). The Intervention Group showed a 1.02 increase in their total ITERS-3 score.

In delving deeper into the coaching intervention, there were some interesting findings that did not reach a statistically significant level because of the small number of programs that participated in the Intervention Group from pretest to posttest. For example, those programs that scored highest on the ITERS-3 post-test utilized prior observations more often (77% vs 49%), the sessions were longer (on the average 10% longer), they were more focused (5%), they were more often planned (20%). When coaching occurred under these circumstances ITERS-3 scores were higher (4.50+). When a regression equation was run with ITERS-3 as the dependent variable, prior observations, length of coaching session, and the number of professional development hours had the greatest beta weights ($\beta > 1.90$). And finally, with both the Comparison Group and the Intervention Group, there is a correlation between the number of professional development module hours taken and increases on the post-test ITERS-3 scores ($r = .33$; $p < .05$). Additional evidence that the Better Kid Care modules are effective at improving the overall quality of care. It will be interesting to compare the results from this study with future studies to determine the expected gains on the ITERS-3 with other Coaching Interventions.

And when correlations were run between the number of professional development module hours and the ITERS-3 Subscales, there were statistically significant relationships with the **Language Subscale** ($r = .48$; $p < .01$) and with the **Program Subscale** ($r = .36$; $p < .05$). Within the Subscales in looking at the specific Items, the following ITERS-3 Items were found to have a statistically significant relationship with the number of professional development module hours**: Item 1 – Indoor Space** ($r = .37$; $p < .05$); **Item 9 – Talking With Children** ($r = .33$; $p < .05$); **Item 10 – Encouraging Vocabulary Development** ($r = .35$; $p < .05$); **Item 11 – Responding to Children's Communication** ($r = .51$; $p < .01$); **Item 12 – Encouraging Children to Communicate** ($r = .37$; $p < .05$); **Item 26 - Supervision** ($r = .36$; $p < .05$); and **Item 31 – Schedule and Transitions** ($r = .36$; $p < .05$).

DISCUSSION

The purpose of this report was to provide an overview to the effectiveness of the online coaching program as implemented by the Better Kid Care Project, provide basic descriptive and demographic analyses from a pre-test data collection effort, and provide information about the coaching intervention in particular. The pre-test analyses equivalency testing was within acceptable ranges when comparing the intervention group and comparison groups on t-tests and One-way ANOVA's.

Post-test analyses demonstrated that both the intervention and comparison groups increased on the ITERS-3 by the same amount, approximately a full 1 point increase. Actually this was not un-expected because the Comparison group received the Better Kid Care online modules which is, in itself, an

innovative training delivery system short of actual online coaching. As demonstrated in the above Results Section, the number of professional development modules taken via the Better Kid Care Online delivery system correlated with increases on the post-test ITERS-3.

The coaching analyses which looked at the focus of coaching, the time, the location, and if observations were done with the coaching prior to the coaching did not show any significant differences; however, it did demonstrate trends showing more time spent in observations and these observations were more focused and goal oriented in those classrooms which had higher ITERS-3 scores.

So, the bottom line is, if you want to increase the quality of care, using either the Online Coaching or the Better Kid Care Online Modules appear from this study to have similar positive results for the programs in increasing their ITERS-3 scores, with the Online Coaching being a tad better at increasing the scores, especially for the Space and Furnishings Sub-Scale. So longer, planned, & focused coaching observations produced the best results related to the ITERS-3.

LIMITATIONS

Sample size would have been sufficient but with the loss of almost 50% of the intervention classrooms, sample size became an issue and impacted the statistically analyses. With a more sufficient sample size, based on the trends in the data, levels of significance would have been attained.

All programs were high performing programs, STAR 3 or 4 as well as scoring 4.93 on average with an earlier version of the ITERS-R, were seasoned programs with a long history of operating in their local communities, and the comparison group was able to take Better Kid Care Online Modules which was a good thing. However, statistically it appears that we started out at a much higher ITERS-3 score level than what would have been generally expected which left less variance in the data for improvement. This led to scale compression in which score increases becoming more difficult. Both the intervention and comparison groups increased at about the same rate from pre to post-test, 1 point on the ITERS-3. It would have been interesting to have a third group which did not get the online coaching nor the online modules offered by the Better Kid Care project but just the run of the mill type of training offered in the Pennsylvania training system.

FUTURE RESEARCH

For future research, the online coaching intervention needs to be utilized with lower performing programs, more at a Star 1 and 2 levels as well as having a comparison group that is truly a control group and does not receive the Better Kid Care online modules. Because the programs selected for this study were such high performers at the upper end of the Pennsylvania QRIS Keystone STARS protocol, there was very little movement that was possible. The programs were starting at a higher level than the average ECE program.

Also, comparing the BKC Online Modules and the Coaching Programs to a group of programs that were receiving the usual run of the mill training would be an interesting comparison. This is where it is hypothesized that highly significant differences would be detected between either the BKC Online Modules or the Coaching Programs when compared to the Control Group.

Note [1] = Legend for Chart 3:

Space 1 = Indoor Space
Space 2 = Furnishing for care play and learning
Space 3 = Room arrangement
Space 4 = Display for children
Personal 5 = Meals/snacks
Personal 6 = Diapering/toileting
Personal 7 = Health practices
Personal 8 = Safety practices
Language 9 = Talking with children
Language 10 = Encouraging vocabulary development
Language 11 = Responding to children's communication
Language 12 = Encouraging children to communicate
Language 13 = Staff use of books with children
Language 14 = Encouraging children's use of books
Activities 15 = Fine motor
Activities 16 = Art
Activities 17 = Music and movement
Activities 18 = Blocks
Activities 19 = Dramatic play
Activities 20 = Nature/science
Activities 21 = Math/number
Activities 23 = Promoting acceptance of diversity
Activities 24 = Gross motor
Interaction 25 = Supervision of gross motor play
Interaction 26 = Supervision
Interaction 27 = Peer interaction
Interaction 28 = Staff-child interaction
Interaction 29 = Providing physical warmth/touch
Interaction 30 = Guiding children's behavior
Program 31 = Schedule and transitions
Program 32 = Free play
Program 33 = Group play activities

Richard Fiene, Ph.D., Senior Research Psychologist, Research Institute for Key Indicators (RIKILLC); and Affiliate Professor, Prevention Research Center, Penn State University. RFiene@RIKInstitute.com or RJF8@psu.edu.

Richard Fiene

6

Using a Statistical-Indicator Methodology for Accreditation

During 1991-92 the National Child Care Association (NCCA), a 4,500-member organization representing the private-for-profit sector in child care in the United States, pilot tested an indicator checklist system as a part of the developing and implementing of the National Early Childhood Program Accreditation system (NECPA). In the development of any indicator checklist system (ICS), a comprehensive set of standards must be used as the backdrop for creating key indicators. The NECPA system used as its set of comprehensive standards the NAEYC accreditation criteria.

NCCA's reason for developing the NECPA system was to streamline accreditation without losing the comprehensiveness of self-study. This was a formidable endeavor in attempting to be cost efficient without compromising effectiveness. NCCA tried an ICS methodology because of its successful use in the development of licensing and monitoring systems. However, the methodology had never been used with accreditation.

NAEYC's accreditation system had a long history of development and revision. It is clearly recognized in the early childhood field as the standard for program quality. Some states have built child care funding standards based upon NAEYC accreditation. Because of its reliance on an instrument-based program monitoring system and its comprehensiveness, NAEYC accreditation fits nicely with the ICS methodology. These aspects of NAEYC's system are two major criteria in the development and use of the key-indicator methodology, and this relationship between an indicator checklist and its respective comprehensive instrument is critical. NECPA is the indicator checklist, with NAEYC being the comprehensive tool. Indicator methodology is always based upon this relationship with the comprehensive tool, which means that the indicator checklist cannot be used solely as the review instrument. In other words, the indicator checklist (NECPA) can never replace the comprehensive instrument (NAEYC).

Key-indicator methodology, an ICS developed in the late 1970s, responded to the Federal Interagency Day Care Requirements' (FIDCR) need to create a cost-effective and efficient monitoring system that would measure compliance with regulations. The ICS is based upon a statistical methodology that has the ability to predict overall compliance with state or federal regulations, based upon a very short list of key regulatory indicators. These key regulatory indicators predict statistically the probability that a program will either be in compliance or out of compliance with state

regulations in the aggregate. If a program is in compliance with the key regulatory indicators, the chances are that the program is in compliance with all regulations. The reverse is also true, if a program is out of compliance with the key regulatory indicators, the chances are that the program is out of compliance with other regulations.

A pilot study conducted between summer 1991 and spring 1992 used the NAEYC self-study process with an abbreviated version of the NAEYC criteria, the new American Academy of Pediatrics/American Public Health Association (AAP/APHA) child care standards, and selected criteria from a national licensing database of key indicators maintained at the Pennsylvania State University at Harrisburg. Trained observers (graduate students in the university's early childhood program) administered the two tools at selected sites throughout Pennsylvania. Results from the observations were tabulated into overall percentage scores on the program as well as component scores by administration, curriculum, physical environment, health and safety, and staffing (Figures 1 and 2). The overall correlations were all statistically significant, showing a very strong relationship (r) between the two tools ($r = .93+$, with component scores = .88 to .96).

These results should not be surprising because the key indicators are essentially a subset of the overall comprehensive instrument used by NAEYC, with only some modifications from the APHA/AAP standards and the National Key Indicator Data Base. This is an important point because there have been misunderstandings regarding the relationship between the two systems. On the surface the NECPA accreditation system appears as a watered-down version of the NAEYC accreditation system. This is not the case. It is true that the NECPA system has fewer criteria; however, the reason is that these criteria are the key indicators that statistically predict compliance with all NAEYC criteria. For the most part, the NECPA and NAEYC systems would accredit the same programs and defer accreditation on the same programs.

Could NAEYC use the key-indicator statistical methodology in its present system? I think the answer is a qualified yes. For programs seeking reaccreditation, with certain criteria not changing, such as no director turnover, little staff turnover, full license in place, no complaints, and so on, the answer would be in the affirmative. The key-indicator statistical methodology would have to be limited in this way. It should not be used in a reaccreditation where substantial changes have occurred. If changes have occurred, it alters and invalidates the key-indicator statistical methodology, which is always based upon a comprehensive review—the NAEYC self-study.

Remember that the real advantage to using a key-indicator methodology is in the time saved by a high-quality program in which there is high compliance with all criteria. Also, there is the time saved in the review process by validators and national accreditation commissions.

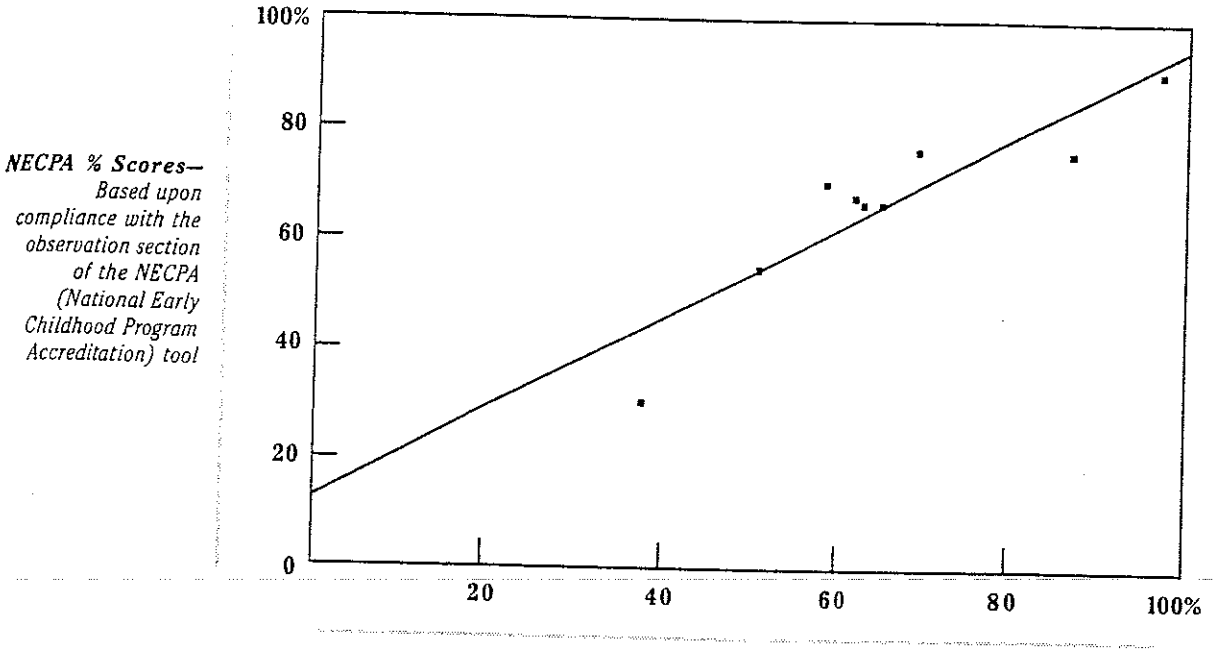
The key-indicator methodology could be used with a sample group of NAEYC-accredited programs from the national database that NAEYC maintains. The key-indicator statistical methodology could be run on this sample of programs. Key indicators would be developed from the results. This is a relatively straightforward exercise that could be done at the Pennsylvania State University at Harrisburg National Center for Early Childhood Program Evaluation. A sampling of key regulatory indicators from the National Key Indicator Data Base follows:

- staff-child ratios
- group size
- staff training
- director qualifications
- lead-teacher qualifications
- children properly immunized
- children supervised at all times
- developmentally appropriate discipline followed

The purpose in developing the NECPA system and in even suggesting the use of the key-indicator methodology within the existing NAEYC accreditation system is not to water down a comprehensive process but to make it more efficient and effective in reaccrediting programs of very high quality. The use of key indicators also helps to streamline a rather labor-intensive, validator and national commission process and review. Another component of the NECPA system is the automated/computer-generated profile of program compliance with standards. The NECPA system was not designed as a competing system but, rather, a system complementary to NAEYC's using the latest technologies in the monitoring of early childhood programs.

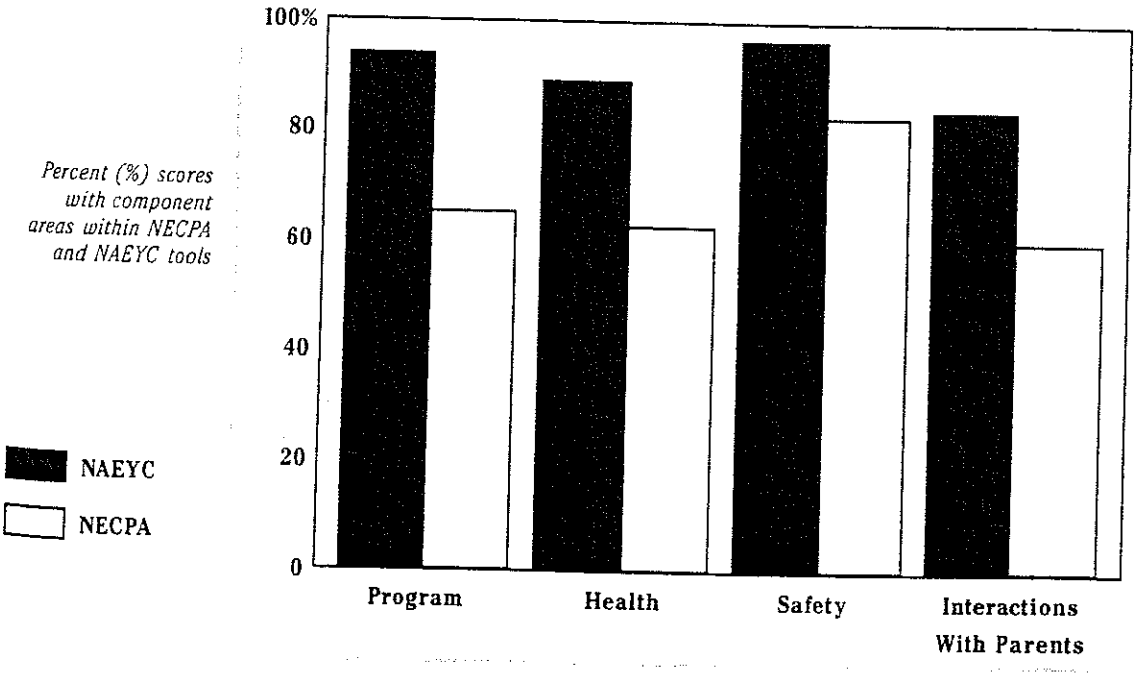
Although the NECPA system was not designed to compete with the NAEYC system, the issue of competition has surfaced, often generating impas-

Figure 1. Relationship of NAEYC and NECPA Accreditation Systems



Source: Used by permission, from R. Fiene, *National Early Childhood Program Accreditation Software System Manual* (Middletown, PA: Early Childhood Program Evaluation Associates, 1992). Copyright © by R. Fiene/NECPA System.

Figure 2. Comparison of NAEYC and NECPA Accreditation System Components



Source: Used by permission, from R. Fiene, *National Early Childhood Program Accreditation (NECPA) Instrument/Manual* (Atlanta, GA: NECPA, Inc., 1992). Copyright © by R. Fiene/NECPA System.

sioned discussion. The accreditation of early childhood programs stands to benefit or suffer, depending on one's perspective. On the positive side, a certain amount of competition generally improves the systems involved; it is to be hoped that both the NECPA and NAEYC systems will improve as a result of their coexistence. However, the negative side is that having two systems is potentially confusing for parents. Although NECPA is a key-indicator approach of the NAEYC accreditation system, this is not clear for the average consumer of child care services. In fact, the key-indicator methodology lacks full acceptance within the early childhood research community. Researchers see the benefit of the key-indicator approach within a licensing system but still question its usefulness within an accreditation system.

With the pilot study and the subsequent research done on the NECPA system, the bottom line is the resulting validation of the NAEYC accreditation system. The NAEYC system remains the comprehensive standard in the field against which all other accreditation systems must be measured. The NECPA results only confirm this, with extremely high correlations between the short version of the NAEYC instrument and the comprehensive version of the NAEYC instrument. The NAEYC system establishes the baseline of program quality for early childhood programs. It is the starting point for programs in their comprehensive self-evaluation. A self-study based upon the key-indicator methodology (NECPA is an example of this approach) can never replace a comprehensive review. The key-indicator self-study can supplement the comprehensive review in subsequent years if there have been few changes since the program conducted its last comprehensive review. It is always assumed that there was very high compliance with the NAEYC criteria in becoming accredited in the first comprehensive self-study.

Future developments in accreditation need to take into account the results of the NECPA pilot and the NAEYC accreditation system. Is the selected use of a key-indicator statistical methodology warranted within the NAEYC system? Only with additional research will we be able to determine the answer.

References

American Academy of Pediatrics, & American Public Health Association (AAP/APHA). 1992. *Caring for our children—National health and safety performance standards: Guidelines for out-of-home child care programs.* Washington, DC: Author.

Fiene, R. 1992. *National Early Childhood Program Accreditation (NECPA) instrument/manual.* Atlanta, GA: NECPA, Inc.

Fiene, R. 1992. *National Early Childhood Program Accreditation software system manual.* Middletown, PA: Early Childhood Program Evaluation Associates.

NAEYC. 1991. *Accreditation criteria & procedures of the National Academy of Early Childhood Programs.* Rev. ed. Washington, DC: Author.

Development of a Regulatory Compliance Scale

Richard Fiene PhD

Penn State's Edna Bennett Pierce Prevention Research Center

April 2025

The purpose of this paper is to provide an alternate paradigm for regulatory compliance measurement in moving from a nominal to an ordinal scale measurement strategy and to introduce a new licensing/regulatory compliance metric: the Regulatory Compliance Scale. Regulatory compliance measurement is dominated by a nominal scale measurement system in which rules are either in compliance or out of compliance. There are no gradients for measurement within the present licensing measurement paradigm. It is very absolute. Either a rule is in full compliance to the letter of the law or the essence of the regulation or it is not. An alternate paradigm borrowing from accreditation and other program quality systems is to establish an ordinal scale measurement system which takes various gradients of compliance into account. With this alternate paradigm, it offers an opportunity to begin to introduce a quality element into the measurement schema. It also allows us to take into consideration both risk and prevalence data which are important in rank ordering specific rules.

So how would this look from a licensing decision making vantage point. Presently, in licensing measurement, licensing decisions are made at the rule level in which each rule is either in or out of compliance in the prevailing paradigm. Licensing summaries with corrective actions are generated from the regulatory compliance review. It is a nominal measurement system being based upon Yes/No responses. The alternate measurement paradigm I am suggesting in this paper is one that is more ordinal in nature where we expand the Yes/No response to include gradients of the particular rule. In the next paragraph, I provide an example of a rule that could be measured in moving from a nominal to ordinal scale measurement schema.

Rather than only measuring a rule in an all or none fashion, this alternate paradigm provides a more relative mode of measurement at an ordinal level. For example, with a professional development or training rule in a particular state which requires, let's say, 6 hours of training for each staff person. Rather than having this only be 6 hours in compliance and anything less than this is out of compliance, let's have this rule be on a relative gradient in which any amount of hours above the 6 hours falls into a program quality level and anything less than the 6 hours falls out of compliance but at a more severe level depending on how far below the 6 hours and how many staff do not meet the requirement (prevalence). Also throw in a specific weight which adds in a risk factor, and we have a paradigm that is more relative rather than absolute in nature.

From a math modeling perspective, the 1 or 0 format for a Yes or No response becomes -2, -1, 0, +1, +2 format. This is more similar to what is used in accreditation systems where 0 equals Compliance and -1 and -2 equals various levels of Non-Compliance in terms of severity and/or prevalence. The +1 and +2 levels equal value added to the Compliance level by introducing a Quality Indicator. This new formatting builds upon the compliance vs non-compliance dichotomy (C/NC) but now adds a quality indicator (QI) element. By adding this quality element, we may be able to eliminate or at least lessen the non-linear relationship between regulatory compliance with rules and program quality scores as measured by the Environmental Rating Scales (ERS) and CLASS which is the essence of the Theory of Regulatory Compliance (TRC). It could potentially make this a more linear relationship by not having the data as skewed as it has been in the past.

By employing this alternate paradigm, it is a first demonstration of the use of the Key Indicator Methodology in both licensing and quality domains. The Key Indicator Methodology has been utilized a great deal in licensing but in few instances in the program quality domain. For example, over the past five years, I have worked with approximately 10 states in designing Licensing Key Indicators but only one state with Quality Key Indicators from their QRIS – Quality Rating and Improvement System. This new paradigm would combine the use in both. It also takes advantage of the full ECPQI2M – Early Childhood Program Quality Improvement and Indicator Model by blending regulatory compliance with program quality standards.

A major implication in moving from a nominal to an ordinal regulatory compliance measurement system is that it presents the possibility of combining licensing and quality rating and improvement systems into one system via the Key Indicator Methodology. By having licensing indicators and now quality indicators that could both be measured by licensing inspectors, there would be no need to have two separate systems but rather one that applies to everyone and becomes mandated rather than voluntary. It could help to balance both effectiveness and efficiency by only including those standards and rules that statistically predict regulatory compliance and quality and balancing risk assessment by adding high risk rules.

I will continue to develop this scale measurement paradigm shift in future papers but wanted to get this idea out to the regulatory administration field for consideration and debate. This will be a very controversial proposal since state regulatory agencies have spent a great deal of resources on developing free standing QRIS which build upon licensing systems. This alternate paradigm builds off the Theory of Regulatory Compliance's key element of relative vs absolute measurement and linear vs non-linear relationships (Fiene, 2022). Look for additional information about this on RIKI Institute Blog - https://rikinstitute.com/blog/.

Introduction to the Regulatory Compliance Scale

The theory of regulatory compliance has been proven in multiple studies over the past four decades and has been utilized extensively in the creation of differential monitoring and its spin off methodologies of risk assessment and key indicators (Fiene, 2025). In fact, differential monitoring would not have been possible without the theory of regulatory compliance because

the paradigm which it replaced, one of one-size-fits-all monitoring or uniform monitoring would have predominated. However, with the theory of regulatory compliance which introduced the importance of substantial regulatory compliance and the search for the right rules/regulations that made a difference in client's lives, rather than emphasizing more or less regulations or rules.

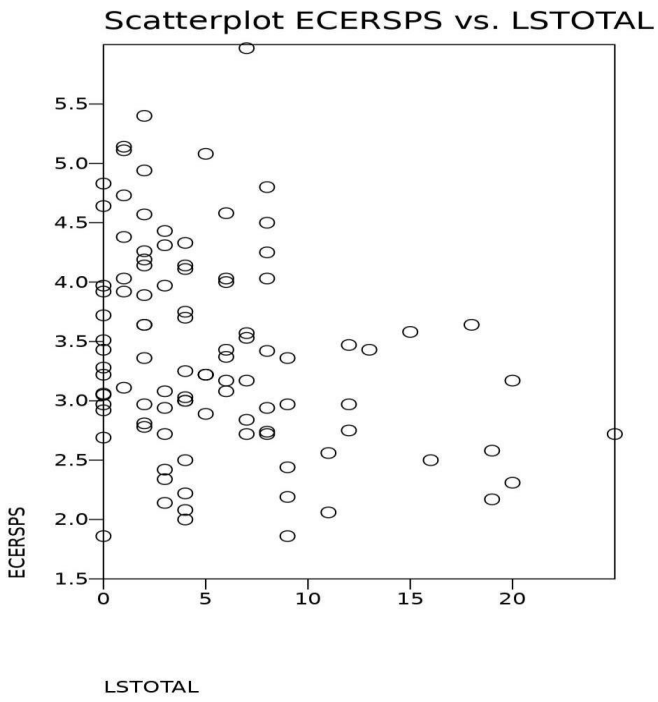
The theory of regulatory compliance has another application when it comes to regulatory compliance measurement in helping to move the licensing field from a nominal based measurement strategy to one of ordinal based measurement. The new measurement strategy is the Regulatory Compliance Scale (RCS) and it is depicted in the following table.

RCS	***Compliance***	***Risk***	***Model***	***Model***
Scale	***Level***	***Level***	***Violations***	***Weights***
7 = A	**Full**	**None**	**0**	**0**
5 = B	**Substantial**	**Low**	**1-3**	**1-3**
3 = C	**Medium**	**Medium**	**4-9**	**4-6**
1 = D	**Low**	**High**	**10+**	**7+**

The above table needs some explanation. The first column is the proposed ordinal scale similar to other scales utilized in the program quality measurement research literature on a 1 – 7 Likert Scale where 7 = Full Regulatory Compliance, 5 = Substantial Regulatory Compliance, 3 = Medium Regulatory Compliance, and 1 = Low Regulatory Compliance. It could also be thought of as an Alpha Scale of A – D as well. The next column has the compliance levels that run from full 100% regulatory compliance to low regulatory compliance. The third column depicts the risk level from none to high which corresponds with the compliance levels. The next two columns depict two models, one unweighted and one in which the rules are weighted with corresponding weights. These models are based upon the two prevailing approaches to rank ordering rules or regulations in the research literature.

The following figures will depict how the scale was conceived based upon empirical evidence in the various studies supporting the theory of regulatory compliance.

The first figure shows the actual individual violation data of the programs compared to their corresponding ECERS scores. There is not a significant relationship between the two as depicted in the graphic.

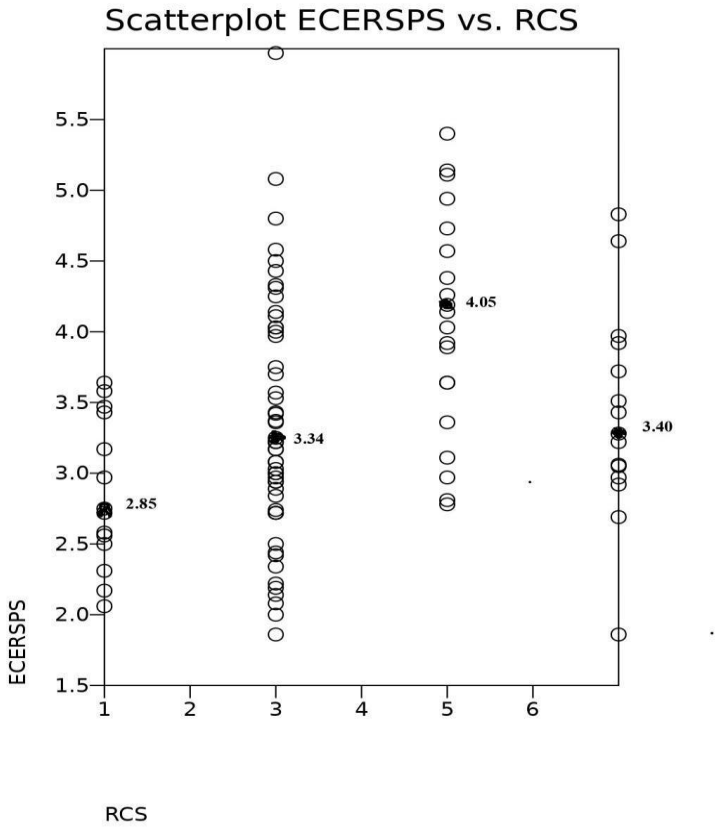


The following figure below depicts what occurs when the individual violation data are grouped according to the theory of regulatory compliance in which a substantial compliance category is introduced, and the data are moved from a nominally based metric to an ordinally based metric of full, substantial, medium, and low regulatory compliance categories. This grouping more clearly reflects the theory of regulatory compliance. It also clearly demonstrates the ceiling effect which is an outcome of the theory of regulatory compliance in which substantial and full regulatory compliance levels are basically equivalent when quality is taken into account. Or at the extreme level which is depicted here where full regulatory compliance quality scores are actually lower than the substantial regulatory compliance quality scores. A footnote about the figures and the scaling: the scales for the first figure are on a lower to higher progression but the higher LSTOTAL represents higher non-compliance where the second figure is also based upon lower to higher but the higher scores represent increased quality and increased regulatory compliance.

So, in reading the change from left to right, these two figures are reversed images of each other. This is just a quirk of the scaling and not a mistake in the plotting of data.

The RCS has been pilot tested in both the non-weighted and weighted models and based upon these studies it appears to be more effective in distinguishing quality amongst the various categories rather than utilizing violation count data. This would be a significant improvement when it comes to licensing measurement. Of course, additional replication studies need to be

completed before it would be recommended as a new Scale to be used for making licensing decisions.



The above figure is dramatically different than the prevailing paradigm which predicts a linear relationship between regulatory compliance and quality which is the paradigm of a uniform monitoring approach. The above results clearly indicate a reconsideration with the introduction of substantial regulatory compliance as an important contributor to overall quality if not the most important contributor to quality. As stated above, these findings have been replicated in several studies conducted over the past several decades.

This would be a major paradigm shift in moving from individual violation data counts to an ordinal scale metric but it does warrant additional research. The problem with individual violation data is that it doesn't take into account the relative risk of the individual rule which could place clients at increased risk of morbidity or mortality. Risk assessment has worked really well when coupled with key indicators in the differential monitoring approach and it appears to be an asset in the development of a Regulatory Compliance Scale (RCS).

Regulatory Compliance Scale Studies

The Regulatory Compliance Scale (RCS) was introduced several years ago and has been used in a couple of validation studies for differential monitoring and regulatory compliance's ceiling effect phenomenon. RCS buckets or thresholds were statistically generated based upon these studies, but it is time to validate those buckets and thresholds to determine if they are really the best model in creating a regulatory compliance scale. Since proposing the RCS, there has been a great deal of interest from jurisdictions in particular from Asian and African nations. Additional statistically based trials were conducted, and this brief report is the compilation of those trials over the past year.

The data used are from several jurisdictions that are part of the international database maintained at the Research Institute for Key Indicators Data Laboratory at Penn State University focusing on program quality scores and rule violation frequency data. These data from the respective databases were recoded into various thresholds to determine the best model. The jurisdictions were all licensing agencies in the US and Canada geographically dispersed where both regulatory compliance and program quality data was obtained from a sample of early care and education programs.

Methodology

The following methodology was used starting with the original RCS buckets/thresholds of Full, Substantial, Medium, and Low regulatory compliance:

RCS Models used for analyses

RCS				**Models**			
		Original	***1***	***2***	***3***	***4***	***5***
	Full	100	100	100	100	100	100
Scaling	***Substantial***	99-98	99-97	99-97	99-98	99-98	99-97
	Medium	97-90	96-90	96-93	97-95	97-85	96-85
	Low	89>	89>	92>	94>	84>	84>

Five alternate models were used to compare the results to the original RCS. The numbers indicate the number of violations subtract from a perfect score of 100. Full regulatory compliance indicates no violations and a score of 100 on the scale. The next bucket of 99-98 indicates that there were 1 or 2 regulatory compliance violations which resulted in a 99-98 score on the scale. This logic continues with each of the models.

The scale score was determined in the following manner: Full Regulatory Compliance = 7; Substantial Regulatory Compliance = 5; Medium Regulatory Compliance = 3; and Low Regulatory Compliance =1. This rubric is how the original RCS scaling was done on a Likert type scale similar to other ECE program quality scales, such as the Environmental Rating Scales.

Results

The following results are correlations amongst the respective RCS Models from Table above compared to the respective jurisdictions program quality tool (Quality1-3): ERS or CLASS Tools.

RCS Model Results compared to Quality Scales

RCS results	Models	Quality1	Quality2	Quality3
Jurisdiction1	**RCS0**	**.26***	**.39***	**.39***
	RCS3	.21	.32*	.33*
	RCS5	.20	.36*	.33*
Jurisdiction2	**RCS0**	**.76****	**.46****	**---**
	RCS3	.12	-.07	---
	RCS5	.18	-.02	---
	RCSF1	.55**	.29*	---
	RCSF2	.63**	.34	---
Jurisdiction3	RCS0	.19	.18	.16
	RCS3	**.21**	**.21**	**.15**
	RCS5	.18	.16	.07
	RCSF1	.17	.17	.10
	RCSF2	.18	.18	.19
Jurisdiction4	RCS0	.24*	---	---
	RCS3	.28*	---	---
	RCS5	**.30***	**---**	**---**
	RCSF1	.21	---	---
	RCSF2	.29*	---	---
Jurisdiction5	RCS0	.06	-.02	.07
	RCS3	.06	-.01	.05
	RCS5	**.08**	**.00**	**.09**
	RCSF1	.00	-.03	.05
	RCSF2	.05	-.03	.05

*Statistically significant .05 level;

**Statistically significant .01 level.

In the above table starting under Jurisdiction2, two new models were introduced based upon the Fibonacci Sequence (Fibonacci1 = RCSF1; Fibonacci2 = RCSF2) and their model structure is in the following Table. The reason for doing this is that the Fibonacci Sequence introduces additional variation into the scaling process.

RCS Fibonacci Models

RCS Fibonacci			**Models**	
		Original	***Fibonacci1***	***Fibonacci2***
	Full	100	100	100
Scaling	***Substantial***	99-98	40	90
	Medium	97-90	20	20
	Low	89>	13	13

A second series of analyses were completed in comparing the RCS models with program quality (Quality1) by running ANOVAs with the RCS models as the independent variable and program quality as the dependent variable. The reason for doing this was the nature of the data distribution in which there was a ceiling effect phenomenon identified which would have had an impact on the correlations in table above. All results are significant at $p < .05$ level with the exception of Jurisdiction2.

ANOVAs Comparing the RCS Models with Program Quality

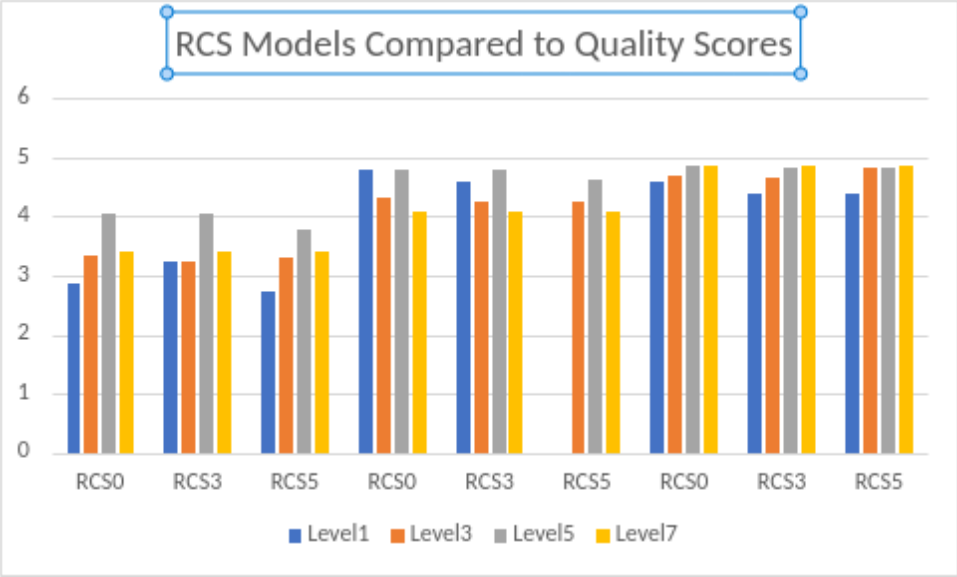
Jurisdictions	**Model**	**Level 1**	**Level 3**	**Level 5**	**Level 7**
Jurisdiction1	**RCS0**	**2.85**	**3.34**	**4.05**	**3.40**
	RCS3	3.24	3.23	4.05	3.40
	RCS5	2.73	3.32	3.77	3.40
Jurisdiction2	RCS0	4.81	4.31	4.80	4.10
	RCS3	**4.59**	**4.25**	**4.80**	**4.10**
	RCS5	---	4.26	4.64	4.10
Jurisdiction3	RCS0	4.59	4.68	4.86	4.87
	RCS3	**4.38**	**4.67**	**4.83**	**4.87**
	RCS5	4.38	4.83	4.83	4.87
Jurisdiction4	RCS0	37.81	37.01	44.28	41.96
	RCS3	36.57	38.60	44.28	41.96
	RCS5	**33.46**	**36.53**	**43.10**	**41.96**
Jurisdiction5	RCS0	3.93	4.17	4.28	4.07
	RCS3	4.02	4.24	4.28	4.07
	RCS5	**3.75**	**4.13**	**4.26**	**4.07**

Discussion

Based upon the above results, it appears that the original RCS model proposed in 2021 is still the best model to be used, although the Fibonacci Sequence model is a close second in some of the jurisdictions. This model will need further exploration in determining its efficacy as a replacement or enhancement to the original RCS Model.

The bottom line is that the original RCS Model is as good as any and no other model is consistently better than all the rest. The RCS Model does have a slight edge over Regulatory Compliance Violation RCV frequency counts in some jurisdictions but not in others. It is much easier to interpret the relationship between quality and the RCS models than it is to interpret the results from the quality scores and the RCV data distribution. So, the recommendation would be for licensing agencies to think about using this new scaling technique in one of its model formats to determine its efficacy. Pairing up RCS and RCV data side by side by licensing agencies would be important studies to determine which approach is the better approach.

The below graphic depicts the relationship between the RCS Models (0, 3, 5) when compared to the quality scores (1-6) clearly showing the ceiling effect and diminishing returns effect phenomenon so typical of regulatory compliance data when compared to program quality. These graphs are from the first three jurisdictions (1, 2, 3) from the above tables.



Additional Analyses Comparing the 11 Studies

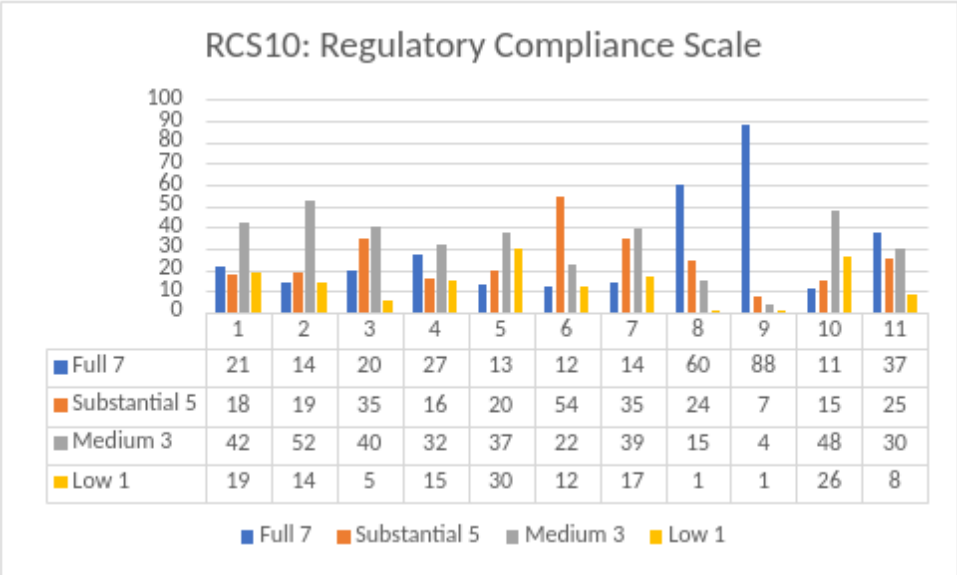
This section provides the results from 11 studies from 10 states and Canadian Provinces in which the proposed new Regulatory Compliance Scale (RCS) was utilized as a byproduct of a differential monitoring implementation or validation study. These studies were undertaken over a decade long period (2013-2023).

The RCS was based upon the following rubric: Full Regulatory Compliance (100%) or no violations = 7; Substantial Regulatory Compliance (99-98) or 1-2 violations = 5; Medium Regulatory Compliance (97-90) or 3-10 violations =3; and Low Regulatory Compliance (89 or less) or 11 or more violations = 1.

These are the results from these 10 jurisdictions which are presented in the following Table (all results are presented as percents of programs that fell into the scaling 1-7). Under the Studies, the number of the specific study is provided, followed by the sample size, followed by if it is in the USA (US) or Canada (CA).

RCS Scale			**RCS Scaling**		
Studies	**7=Full**	**5=Substantial**	**3=Medium**	**1=Low**	**Comments**
1-403-US	21%	18%	*42%*	19%	***High Med NC***
2-104-US	14%	19%	*52%*	14%	***High Med NC***
3-422-US	***20%***	***35%***	*40%*	***5%***	***OK***
4-219-CA	***27%***	***16%***	*32%*	***15%***	***OK***
5-60-CA	13%	20%	*37%*	***30%***	***High NC/Low C***
6-585-US	***12%***	***54%***	*22%*	***12%***	***OK***
7-255-US	***14%***	***35%***	*39%*	***17%***	***OK***
8-1399-US	***60%***	24%	15%	***1%***	***Low NC/High C***
9-2116-US	***88%***	7%	4%	***1%***	***Low NC/High C***
10-482-US	11%	15%	*48%*	*26%*	***High NC/Low C***
11-3070-US	***37%***	***25%***	*30%*	***8%***	***OK***

In looking at the results, it is preferrable to have most of the programs at either a full or substantial regulatory compliance level (7 or 5) and to have fewer programs at the medium or low regulatory compliance level (3 or 1). But in those jurisdictions where there are higher percentages of programs at the medium or low levels of regulatory compliance, it could be that their enforcement of rules and regulations is more stringent. This potential result needs further investigation to get to the root cause of these differences because there is a good deal of variation across the jurisdictions as is evident from the above table.



	1	2	3	4	5	6	7	8	9	10	11
Full 7	21	14	20	27	13	12	14	60	88	11	37
Substantial 5	18	19	35	16	20	54	35	24	7	15	25
Medium 3	42	52	40	32	37	22	39	15	4	48	30
Low 1	19	14	5	15	30	12	17	1	1	26	8

Based upon the above studies and results, the regulatory compliance scale (Fiene, 2022) which appears from recent studies to be a better metric in measuring regulatory compliance than just counting the number of violations that a program has related to their respective rules, regulations, or standards. So how does the regulatory compliance scale work. It essentially puts violations into buckets of regulatory compliance as follows: full compliance (100%) or no violations; substantial compliance (99-98%) or 1-2 violations; mediocre compliance (97-90%) or 3-9 violations; and lastly low/non-optimal compliance (89% or lower) or 10+ violations. Why buckets, because logically it works, it is the way we think about regulatory compliance. It is a discrete rather than continuous metric and logically fits into these four categories. This is based upon 50 years of research into regulatory compliance data distributions and when the data are moved from frequency counts of violation data into these buckets/categories, the math works very well in identifying the better performing programs.

Regulatory Compliance Scale Extensions

Depicted below is a regulatory compliance grid model showing the relationship between regulatory compliance (RC) and program quality (PQ).

An explanation of the below chart will demonstrate how regulatory compliance and program quality in human service facilities interact. The horizontal blue axis depicts the various levels of regulatory compliance while the vertical green axis depicts the various levels of program quality of facilities. It ranges from 1-5 or low to high for each axis. The red "**X's**" represents the relationship that has been identified in the research literature based upon the theory of regulatory compliance in which there is either a plateau effect or a downturn in quality as regulatory compliance increases. The one italicized "***X***" is an outlier that has also been identified in the research literature in which some (it does not happen often) low compliant programs really are at a high-quality level.

It is proposed in order to mitigate the plateau effect with regulatory compliance and program quality standards because regulatory compliance data distributions are severely skewed which means that many programs that have questionable quality are being included in the full (100%) compliance domain. When regulatory compliance standards are increased in their quality components this will lead to a higher level of overall quality as depicted in the "**XX**" cell all the way on the lower right. It also helps to mi gate the severe skewness in the regulatory compliance data distribution. The data distribution does not approximate a normally distributed curve which is the case with the program quality data distribution.

Regulatory Compliance x Program Quality Grid Model

PQ/RC ->	1 Low	2 Med	3 Substantial	4 Full 100%	5QualityAdditions
1 Low	XXX				
2		XX			
3 Med			XX	XXX	
4			XX	X	
5 High	*X*				XX

By utilizing this model, it helps to deal more directly in taking a non-linear relationship and making it linear again when comparing regulatory compliance with program quality. This model provides a theoretical approach supporting what many state licensing administrators are thinking from a policy standpoint: add more quality to health and safety rules/regulations. This grid/matrix also depicts the three regulatory compliance models: Linear, Non-linear, and Stepped.

Here is another potential extension of the Regulatory Compliance Scale using the ECPQIM DB – Early Childhood Program Quality Improvement and Indicator Model Data Base, it is possible to propose developing and using a Regulatory Compliance Scoring System and Scale (RC3S). This new proposed RC3S could be used by state human service agencies to grade facilities as is done in the restaurant arena. Presently, in the human service field, licenses are issued with a Certificate of Compliance but generally it does not indicate what the regulatory compliance level is at. This new proposal would alleviate this problem by providing a scale for depicting the level of regulatory compliance.

The ECPQIM DB is an international data base consisting of a myriad group of data sets drawn from around the USA and Canada. It has been in the making over 40 years as of this writing, so its stability and generalizability have been demonstrated. What follows is the chart depicting the RC3S.

Regulatory Compliance Scoring System and Scale (RC3S)

Color	Non-Compliance Level	Regulatory Compliance Level
Blue	**0**	**Full Compliance**
Green	**1-2**	**Substantial Compliance**
Yellow	**3-6**	**Mid-Range Compliance**
Orange	**7-9**	**Low Compliance**
Red	**10-15+**	**Very Low Compliance**

It is evident from the above chart that the color go from blue to red which indicates an increasing risk of non-compliance and a lower level of overall regulatory compliance, which is not a good thing in the licensing field. Non-compliance levels indicate the number of rules or regulations or standards that are not complied with. And lastly, the regulatory compliance level indicates the movement from full (100% regulatory compliance with all rules) to very low compliance with rules. These ranges for the scaling are based on 40 years of research in understanding and plotting the data distributions around the world related to regulatory compliance in the human services. These results have consistently appeared over this 4-decade time period and show no signs of changing at this point.

Regulatory Compliance Scaling for Decision Making

There is a lack of empirical demonstrations of regulatory compliance decision making. In the past, I have used the methodologies of key indicators, risk assessment and the resultant differential monitoring techniques of how often and what should be reviewed for decision making. What has not been addressed is decision making based upon comprehensive reviews when all regulations are assessed. This section addresses how empirical evidence taken from the past 40+ years of establishing and researching a national database for regulatory compliance can help lead us to a new scaling of regulatory compliance decision making.

In analyzing regulatory compliance data, it becomes perfectly clear that the data have very little variance and are terribly skewed in which the majority of programs are in either full or substantial compliance with all the respective regulations. Only a small handful of programs fall into the category of being in low compliance with all the regulations.

The proposed scaling has three major decision points attached to regulatory compliance scores. Either programs are in full or substantial compliance, in low compliance or somewhere in the middle. Full or substantial regulatory compliance is 100% or 99-98% in regulatory compliance. Low regulatory compliance is less than 90% and mid-regulatory compliance is between 97%-90%. These ranges may seem exceptionally ght but based upon the national database on regulatory compliance that I maintain at the Research Institute for Key Indicators (RIKILLC) these are the ranges that have formed over the past 40 years. These data ranges should not come as a surprise because we are talking about regulatory compliance with health and safety standards. These are not quality standards; these are basic protections for clients. The data are not normally distributed, not even close as is found in quality tools and standards.

What would a **Regulatory Compliance Decision-Making Scale** look like:

Data	Level	Decision
100-98%	***Full/Substantial***	***License***
97-90%	***Mid-Range***	***Provisional***
89% or less	***Low***	***No-License***

States/Provinces/Jurisdictions may want to adjust these levels, and the scaling based upon their actual data distribution. For example, I have found certain jurisdictions to have very unusually skewed data distributions which means that these ranges need to be ghten even more. If the data distribution is not as skewed as the above scale, then these ranges may need to be more forgiving.

This regulatory compliance decision making scale does not take into account if abbreviated methodologies are used, such as risk assessment or key indicator models that are used in a differential monitoring approach. The above scale is to be used if a jurisdiction decides not to use a differential monitoring approach and wants to measure regulatory compliance with all regulations and complete comprehensive reviews.

Conclusion

The Theory of Regulatory Compliance (Fiene, 2019) and bringing substantial compliance to the fore front of regulatory science has been written about a great deal. This paper builds upon these previous assertions and expands them into some practical applications that can be utilized within regulatory science as it relates to licensing measurement, regulatory compliance scaling, and monitoring systems paradigms. This paper has introduced the Regulatory Compliance Scale which is a departure in how best to measure regulatory compliance. This new scale along with the proposed Uncertainty-Certainty Matrix (Fiene, 2025b) provides a robust licensing measurement and program monitoring strategy. This paper provides the last piece of a differential monitoring approach that includes instrument-based program monitoring, key indicators, risk assessment, and the uncertainty-certainty matrix.

Regulatory Compliance has been always approached as an all or none phenomenon, whether a rule is in compliance, or it is not. There is no in-between or shades of gray or partial compliance. This worked when the prevailing paradigm was that full regulatory compliance and program quality were a linear relationship. This was the assumption but not empirically verified until the later 1970's-1980's. When this assumption was put to an empirical test, it did not hold up but rather a curvilinear relationship between regulatory compliance and program quality was discovered. This upset the prevailing paradigm and suggested we needed a new approach to addressing the relationship between regulatory compliance and program quality.

It became clear after these findings in the 1970's-80's and then in the 2010's when replication studies were completed that substantial regulatory compliance could not be ignored based upon this new theory of regulatory compliance in which substantial compliance acted as a "sweet spot" of best outcomes or results when comparing regulatory compliance and program quality scores. The nominal metric needed to be revised and more of an ordinal metric was to be its

replacement. Because now it wasn't just being in or out of compliance, but it mattered which rules were in or out of compliance and how they were distributed. This revised application involved aggregate rules and does not apply to individual rule scoring. The studies completed between 1970 and 2010 involved aggregate rules and not individual rules. To determine if the nominal to ordinal metric needs to be revised still needs empirical data to back this change.

The introduction of substantial compliance into the regulatory compliance measurement strategy moved the field from an instrument-based program monitoring into a more differential monitoring approach. With differential monitoring this approach considered which rules and how often reviews should be done. Also, a new Regulatory Compliance Scale was proposed to take into account the importance of substantial compliance based upon the regulatory compliance theory of diminishing returns. As this Regulatory Compliance Scale has evolved within the licensing health and safety field it needs further revision in which program quality can be infused into the decision making related to individual rules. Remember that the original studies were concerned about rules in the aggregate and not individual rules. It has now become apparent that in dealing with the infusion of quality into rule formulation, a return to the individual rule approach makes the most sense.

The next iteration of the Regulatory Compliance Scale will contain the following categories: Exceeding full compliance, Full compliance, Substantial compliance, and Mediocre compliance to adjust for the infusion of the quality element. This differs slightly from the original aggregate rule Regulatory Compliance Scale where the categories were Full compliance, Substantial compliance, Mediocre compliance and Low compliance where only licensing health and safety elements were considered (see the Table below which depicts the regulatory compliance scales and program monitoring systems side by side).

Without the Theory of Regulatory Compliance, differential and integrative monitoring would not be needed because regulatory compliance would have had a linear relationship with program quality and full compliance would have been the ultimate goal. There would have been no need for targeted rule enforcement or reviews because all rules would have had an equal weight when it came to protecting clients and any individual rule would have predicted overall compliance. But it "just ain't so" as it is said. The need to make adjustments is brought about by the theory and it has not been the same ever since.

Regulatory Compliance Scales and Program Monitoring Systems

Scoring Level	**Individual Rule**		**Aggregate Rules**	**Individual Rule**
Scale	**Instrument based**	***Scale***	**Differential**	**Integrated**
7	Full Compliance	7	Full Compliance	Exceeds Compliance
-	---	5	Substantial	Full Compliance
-	---	3	Mediocre	Substantial
1	Out of Compliance	1	Low	Mediocre/Low

The above table attempts to summarize in tabular form the previous paragraphs in describing the relationship between program monitoring and licensing measurement scaling via a proposed regulatory compliance scale. As one can see this moves the paradigm from a nominal to an ordinal measurement rubric and depicts the differences in the measurement focus either at the

individual rule or aggregate rules scoring levels. It also considers the significance of substantial compliance given the theory of regulatory compliance in which substantial compliance focus is a "*sweet spot*" phenomenon as identified in the regulatory science research literature. It is hoped that the regulatory science field takes these paradigm shifts into consideration in moving forward with building licensing decision making systems and how licenses are issued to facilities.

As a final footnote, keep in mind that the Theory of Regulatory Compliance applies to the relationship between regulatory compliance and program quality and does not apply to regulatory compliance in and of itself related to health and safety. When dealing with regulatory compliance, full compliance is the ultimate goal with individual rules and in determining which rules are predictive rules. It is the preferred methodology in order to eliminate false negatives and decreasing false positives in making licensing decisions related to regulatory compliance.

These above concepts all relate to the field of regulatory compliance and how to make informed decisions about licensing, particularly in the context of program monitoring. Here's how they connect:

Regulatory Compliance Scales:

These scales move away from a binary "compliant" or "non-compliant" approach to regulations. Instead, they acknowledge degrees of compliance, recognizing that minor deviations may not be as detrimental as major ones.

They provide a framework for evaluating the severity and frequency of non-compliance, allowing for more nuanced licensing decisions.

Instrument Based Program Monitoring (IBPM):

This is the traditional method of monitoring compliance, relying on standardized instruments and checklists to assess adherence to specific rules.

It's a comprehensive approach, but can be time-consuming and inflexible, potentially leading to over-regulation or missing important aspects of program quality.

Differential Monitoring (DM):

This approach takes into account the risk associated with different regulations, focusing monitoring efforts on areas with the highest potential for harm or non-compliance.

It allows for a more efficient use of resources and can be tailored to the specific needs of each program.

DM often utilizes Regulatory Compliance Scales to determine the severity of non-compliance and guide the level of monitoring needed.

Integrative Monitoring Systems (IMS):

These systems go beyond simply checking compliance and aim to assess the overall quality of a program.

They integrate data from various sources, including IBPM, DM, and other program-specific metrics, to provide a holistic picture of performance.

IMS can inform licensing decisions by considering not only compliance but also program effectiveness in achieving its goals.

Here's a simplified analogy to illustrate the relationships:

Think of regulations as traffic rules.

IBPM is like a police officer checking every car for every violation, regardless of severity.

DM is like a police officer focusing on patrolling areas with high accident rates or known reckless drivers.

Regulatory Compliance Scales are like different levels of fines based on the severity of the traffic violation.

IMS is like a traffic management system that collects data on accidents, traffic flow, and road conditions to optimize traffic flow and safety.

Relationships:

RCS forms the foundation for DM and IMS by providing a way to assess degrees of compliance.

IBPM provides data for RCS and can be incorporated (with adaptations) into DM and IMS.

DM builds on RCS and IBPM by differentiating the intensity of monitoring based on risk and compliance.

IMS is the most comprehensive approach, integrating RCS, IBPM, DM, and additional data sources for a deeper understanding of program performance.

Regulatory Compliance Scales can be used within any of the monitoring approaches to provide a more nuanced assessment of compliance.

IBPM can be a starting point for differential monitoring, providing data on rule compliance to inform risk assessments.

Differential monitoring can be integrated into an integrative monitoring system, along with other data sources, to provide a comprehensive picture of program performance.

Here are some additional points to consider:

The choice of the most appropriate approach will depend on the specific context, such as the type of program being regulated and the available resources.

Implementation of these alternative paradigms requires careful planning and training of regulators and program providers.

Ongoing research and evaluation are needed to refine these approaches and ensure their effectiveness.

These alternative paradigms offer a more flexible and effective approach to licensing decisionmaking compared to the traditional IBPM approach. They allow for a better

understanding of program strengths and weaknesses, optimize resource allocation, and ultimately lead to better regulatory outcomes.

These concepts offer a shift from traditional "one-size-fits-all" compliance models to more flexible and nuanced approaches that consider risk, program quality, and degrees of compliance. This can lead to more efficient and effective regulatory systems that support program improvement while protecting public safety.

Ultimately, these concepts offer alternative paradigms for licensing decision-making, moving away from a rigid "one-size-fits-all" approach to a more nuanced and risk-based system that considers both compliance and program quality.

References:

Fiene, R. (2019). A treatise on Regulatory Compliance. ***Journal of Regulatory Science, Volume 7***, 2019. https://doi.org/10.21423/jrs-v07fiene

Fiene (2022a). Regulatory Compliance Monitoring Paradigms and the Relationship of Regulatory Compliance/Licensing with Program Quality: A Policy Commentary. (2022). ***Journal of Regulatory Science, 10*(1). https://doi.org/10.21423/JRS-V10A239**

Fiene (2022b). Regulatory Compliance Scale, ***RIKINotes Blog***, January 2022.

Fiene (2023a). ***Licensing Measurement & Monitoring Systems***, Research Institute for Key Indicators, Elizabethtown, Pennsylvania.

Fiene (2023b). Ceiling Effect/Diminishing Returns, Regulatory Compliance Scale, and Quality Indicators Scale, ***Mendeley Data,*** doi: 10.17632/gc423hprcs.1

Fiene (2025a). Finding the Right Rules. ***American Scientist, Volume 113, 1***. pps 16-19.

Fiene (2025b). The Uncertainty-Certainty Matrix for Licensing Decision Making, Validation, Reliability, and Differential Monitoring Studies, ***Knowledge Journal***, under review.

NARA (2023). ***Saskatchewan Differential Monitoring/Quality Indicators Scale Validation Study***, National Association for Regulatory Administration, Fredericksburg, Virginia.

Richard Fiene PhD, Senior Research Psychologist, Research Institute for Key Indicators (RIKILLC); Professor of Psychology (ret), Penn State University; Senior Research Consultant, National Association for Regulatory Administration (NARA). Email: Rfiene@RIKInstitute.com Website: http://RIKInstitute.com